



**TUGAS AKHIR - SS141501**

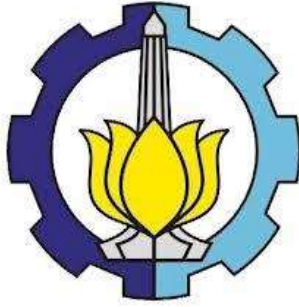
**DIAGNOSA KEJANG EPILEPSI BERDASARKAN SINYAL  
EEG MENGGUNAKAN METODE *K-NEAREST NEIGHBOR***

**CARIANTO HOTTUA SIHOMBING  
NRP 062114 4000 0102**

**Dosen Pembimbing  
Dr. Santi Wulan Purnami, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**





**TUGAS AKHIR - SS141501**

**DIAGNOSA KEJANG EPILEPSI BERDASARKAN SINYAL  
EEG MENGGUNAKAN METODE *K-NEAREST NEIGHBOR***

**CARIANTO HOTTUA SIHOMBING  
NRP 062114 4000 0102**

**Dosen Pembimbing  
Dr. Santi Wulan Purnami, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**





**FINAL PROJECT - SS 141501**

# **EPILEPTIC SEIZURE DETECTION BASED ON EEG SIGNAL USING K-NEAREST NEIGHBOR METHOD**

**CARIANTO HOTTUA SIHOMBING  
SN 062114 4000 0102**

**Supervisor  
Dr. Santi Wulan Purnami, M.Si**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**



## LEMBAR PENGESAHAN

### DIAGNOSA KEJANG EPILEPSI BERDASARKAN SINYAL EEG MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

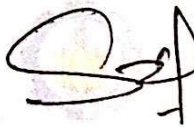
#### TUGAS AKHIR

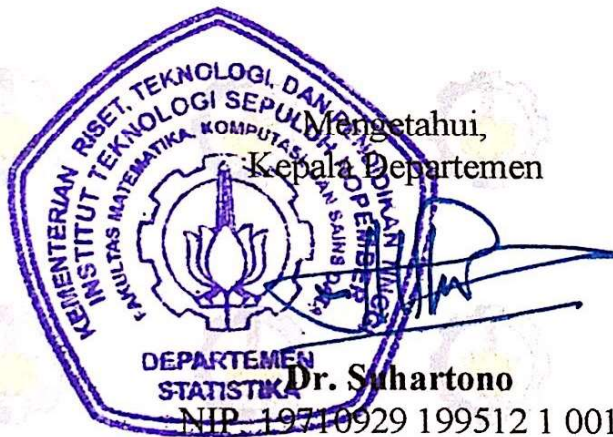
Diajukan untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Sains  
pada  
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :  
**Carianto Hottua Sihombing**  
NRP. 062114 4000 0102

Disetujui oleh Pembimbing :

**Dr. Santi Wulan Purnami, M.Si**  
NIP. 19720923 199803 2 001

(  )



SURABAYA, JULI 2018





# **DIAGNOSA KEJANG EPILEPSI BERDASARKAN SINYAL EEG MENGGUNAKAN METODE *K-NEAREST NEIGHBOR***

**Nama Mahasiswa : Cariato Hottua Sihombing**  
**NRP : 062114 4000 0102**  
**Departemen : Statistika-FMKSD-ITS**  
**Dosen Pembimbing : Dr. Santi Wulan Purnami, M.Si**

## **Abstrak**

*Epilepsi merupakan salah satu kelainan neurologi yang paling umum terjadi. Penyakit epilepsi adalah suatu kondisi yang dapat menjadikan seseorang mengalami kejang secara tiba-tiba dan berulang. Kejang adalah kejadian epilepsi dan merupakan ciri epilepsi yang harus ada, tetapi tidak semua kejang merupakan manifestasi epilepsi. Analisis Electroencephalography (EEG) merupakan metode utama dalam mendiagnosa kejang epilepsi. Sinyal EEG memiliki distribusi non-gaussian, non-stasioner dan memiliki sifat non-linier sehingga diperlukan metode komputerasi dalam mendiagnosa kejang epilepsi. EEG terdiri atas lima bagian frekuensi yaitu delta, theta, alpha, beta, dan gamma. Hanya theta, alpha, dan beta yang membawa informasi kejang. Pada penelitian ini metode IIR butterworth filter digunakan untuk memfilter sinyal menjadi gelombang theta, alpha, dan beta. Selanjutnya dari masing-masing sub-frekuensi diekstrak fitur maksimum, minimum, variance, energi, dan entropy yang akan digunakan sebagai variabel prediktor dalam klasifikasi menggunakan metode KNN. Pada langkah analisis klasifikasi dengan menggunakan metode KNN didapatkan nilai K terbaik adalah saat  $K=3$ , selanjutnya dengan menggunakan metode 3-NN didapatkan nilai akurasi, sensitifitas, dan spesitivitas masing-masing sebesar 98,4%, 98,75%, dan 97%.*

**Kata Kunci: *Electroencephalography, Epilepsi, IIR Butterworth, K-Nearest Neighbor***

*(Halaman ini sengaja dikosongkan)*

## **EPILEPTIC SEIZURE DETECTION BASED ON EEG SIGNAL USING *K-NEAREST NEIGHBOR* METHOD**

**Name** : Carianto Hottua Sihombing  
**Student Number** : 062114 4000 0102  
**Department** : Statistics-FMKSD-ITS  
**Supervisor** : Dr. Santi Wulan Purnami, M.Si

### **Abstract**

*Epilepsy is one of the most common neurological disorders. Epilepsy is a condition that can cause a person to experience seizures suddenly and repeatedly. Seizures are epilepsy events and are characteristic of epilepsy that must exist, but not all seizures are epilepsy manifestations. Electroencephalography (EEG) analysis is the main method to diagnose epileptic seizures. EEG signals have non-gaussian, non-stationary and non-linear distributions so the computerized methods are required to diagnose epileptic seizures. EEG consists of five different frequency bands namely delta, theta, alpha, beta, and gamma. Only theta, alpha, and beta carry seizure information. In this research the IIR butterworth filter method is used to filter the signal into theta, alpha, and beta sub bands. Feature such as maximum, minimum, variance, energy, and entropy is extracted from each of the three sub bands and used as predictor variables in the classification using KNN method. In the step of classification analysis using KNN method, the best value of K is given when  $K = 3$ , then by using 3-NN method, the classification performance values such as accuracy, sensitivity, and specificity are 98,4%, 98,75% and 97%.*

**Keywords:** *Electroencephalography, Epilepsy, IIR Butterworth, K-Nearest Neighbor*

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa, atas segala berkat dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Diagnosa Kejang Epilepsi Berdasarkan Sinyal EEG Menggunakan Metode *K-Nearest Neighbor*”** dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada :

1. Dr. Santi Wulan Purnami, S.Si, M.Si selaku dosen pembimbing Tugas Akhir dalam memberikan bimbingan selama penyusunan Tugas Akhir.
2. Dr. Purhadi, M.Sc. dan Dr. Irhamah, S.Si, M.Si. selaku dosen penguji yang telah banyak memberi masukan kepada penulis dalam penyelesaian Tugas Akhir.
3. Dr. Suhartono selaku ketua Departemen Statistika dan Bapak Dr. Sutikno, M.Si. selaku Ketua Program Studi S1 Statistika ITS. Serta Dra. Madu Ratna, M.Si. selaku dosen wali atas dukungan yang diberikan selama proses perkuliahan.
4. Seluruh dosen dan *staff* pengajar Program Studi Sarjana Departemen Statistika FMKSD ITS yang telah membantu penulis selama proses perkuliahan.
5. Kedua orang tua, Kakak, Abang, Serta Adik yang telah memberikan dukungan dan doa dalam menyelesaikan Tugas Akhir ini.
6. Teman-teman seperjuangan EEG Bayu, Dhamai, dan Nuruddin yang telah membantu dan memberikan masukan dalam penulisan Tugas Akhir ini.
7. Teman-teman kos garuda yang telah menyediakan tempat dan memberikan masukan bagi penulis selama pengerjaan Tugas Akhir.
8. Teman-teman Statistika ITS angkatan 2014, Respect, yang selalu memberikan dukungan kepada penulis selama ini.
9. Semua pihak yang turut membantu dalam pelaksanaan Tugas Akhir yang tidak bisa penulis sebutkan satu persatu.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, Juli 2018

Penulis

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>COVER PAGE</b> .....	iii
<b>LEMBAR PENGESAHAN</b> .....	v
<b>ABSTRAK</b> .....	vii
<b>ABSTRACT</b> .....	ix
<b>KATA PENGANTAR</b> .....	xi
<b>DAFTAR ISI</b> .....	xiii
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR TABEL</b> .....	xvii
<b>DAFTAR LAMPIRAN</b> .....	xix
<b>DAFTAR NOTASI</b> .....	xxi
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan .....	4
1.4 Manfaat .....	5
<b>BAB II INJAUAN PUSTAKA</b> .....	7
2.1 <i>Infinite Impulse Response</i> (IIR) Filter .....	7
2.2 Ekstraksi Fitur .....	10
2.3 <i>K-Nearest Neighbor</i> (KNN) .....	11
2.4 <i>Stratified K-Fold Cross-Validation</i> .....	14
2.5 Ukuran Kinerja Klasifikasi .....	14
2.6 Epilepsi .....	16
2.7 <i>Electroencephalography</i> (EEG) .....	17
<b>BAB III METODOLOGI PENELITIAN</b> .....	19
3.1 Sumber Data .....	19
3.2 Struktur Data .....	19
3.3 Langkah Analisis .....	22
<b>BAB IV ANALISIS DAN PEMBAHASAN</b> .....	25
4.1 Deskripsi Dataset .....	25
4.2 <i>Bandpass</i> Filter Potongan Sinyal .....	27
4.3 Ekstraksi Fitur Sinyal .....	30
4.3.1 Fitur Maksimum .....	30

4.3.2	Fitur Minimum.....	33
4.3.3	Fitur <i>Variance</i> .....	35
4.3.4	Fitur Energi.....	37
4.3.5	Fitur <i>Entropy</i> .....	39
4.4	Pembagian Data <i>Traning</i> dan Data <i>Testing</i> .....	41
4.5	<i>K-Nearest Neighbor</i> .....	42
<b>BAB V KESIMPULAN DAN SARAN</b> .....		49
5.1	Kesimpulan .....	49
5.2	Saran.....	49
<b>DAFTAR PUSTAKA</b> .....		50
<b>LAMPIRAN</b> .....		55
<b>BIODATA PENULIS</b> .....		73



## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Diagram blok <i>Filter</i> IIR .....	8
<b>Gambar 2.2</b> Tipe <i>Magnitude Respon</i> dari Konfigurasi <i>Filter</i> <i>butterworth</i> (a), <i>Chebyshev</i> (b dan c), dan <i>Elliptical</i> (d).....	9
<b>Gambar 3.1</b> Pemasangan Elektoda dalam Perekaman EEG .....	21
<b>Gambar 3.2</b> Diagram Alir Penelitian .....	23
<b>Gambar 4.1</b> Persentase Antara Sinyal Kejang dan Sinyal Non-Kejang.....	25
<b>Gambar 4.2</b> Plot <i>Time Series Data</i> (a) Potongan F1, (b) Potongan N1, (c) Potongan O1, (d) Potongan S1, dan (e) Potongan Z1 .....	26
<b>Gambar 4.3</b> Frekuensi Respon dari Masing-Masing Gelombang .....	28
<b>Gambar 4.4</b> Hasil Filterisasi Sinyal dengan Gelombang (a) Theta, (b) Alpha, dan (c) Beta .....	29
<b>Gambar 4.5</b> Plot Nilai Maksimum Gelombang <i>Theta</i> (a), <i>Alpha</i> (b), dan <i>Beta</i> (c) .....	32
<b>Gambar 4.6</b> Plot Nilai Minimum Gelombang <i>Theta</i> (a), <i>Alpha</i> (b), dan <i>Beta</i> (c) .....	34
<b>Gambar 4.7</b> Plot Nilai <i>Variance</i> Gelombang <i>Theta</i> (a), <i>Alpha</i> (b), dan <i>Beta</i> (c) .....	36
<b>Gambar 4.8</b> Plot Nilai Energi Gelombang <i>Theta</i> (a), <i>Alpha</i> (b), dan <i>Beta</i> (c) .....	38
<b>Gambar 4.9</b> Plot Nilai <i>Entropy</i> Gelombang <i>Theta</i> (a), <i>Alpha</i> (b), dan <i>Beta</i> (c) .....	40
<b>Gambar 4.10</b> Pembagian Data Dengan Menggunakan <i>Stratified 5-Fold Cross Validation</i> .....	41

<b>Gambar 4.11</b> Ilustrasi Penentuan Kelas Berdasarkan	
Mayoritas Kelas Jarak Terdekat.....	44
<b>Gambar 4.12</b> Grafik Perbandingan Performansi Nilai k.....	45

## DAFTAR TABEL

	Halaman
<b>Tabel 2.1</b> <i>Confusion Matrix</i> .....	15
<b>Tabel 3.1</b> Struktur Data Sinyal Tahapan Filterisasi .....	20
<b>Tabel 3.2</b> Variabel Tahapan Klasifikasi.....	21
<b>Tabel 3.3</b> Struktur Data Tahapan Klasifikasi .....	22
<b>Tabel 4.1</b> Rentang Frekuensi Gelombang <i>Tetha</i> , <i>Alpha</i> , dan <i>Beta</i> yang Digunakan Dalam Penelitian.....	28
<b>Tabel 4.2</b> Hasil Fitur Maksimum Masing-Masing <i>Sub-band</i> ...	31
<b>Tabel 4.3</b> Hasil Fitur Minmum Masing-Masing <i>Sub-band</i> .....	33
<b>Tabel 4.4</b> Hasil Fitur <i>Variance</i> Masing-Masing <i>Sub-band</i> .....	35
<b>Tabel 4.5</b> Hasil Fitur Energi Masing-Masing <i>Sub-band</i> .....	37
<b>Tabel 4.6</b> Hasil Fitur <i>Entropy</i> Masing-Masing <i>Sub-band</i> .....	39
<b>Tabel 4.7</b> Ilustrasi Sampel Data .....	42
<b>Tabel 4.8</b> Jarak <i>Euclidean</i> Data ke-6 dengan Data Lainnya....	43
<b>Tabel 4.9</b> Perbandingan Performansi Klasifikasi Dengan Nilai K yang Berbeda .....	44
<b>Tabel 4.10</b> <i>Confusion Matrix</i> Klasifikasi Terbaik dengan Metode 3-NN.....	46
<b>Tabel 4.11</b> Perbandingan Performansi Klasifikasi dengan k=3 Pada Masing-Masing <i>Fold</i> .....	48

*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1.</b> Data Sinyal EEG untuk Deskripsi Dataset dan Filterisasi Sinyal .....	55
<b>Lampiran 2.</b> Data Hasil Filterisasi Sinyal EEG .....	56
<b>Lampiran 3.</b> Data Hasil Ekstraksi Fitur .....	57
<b>Lampiran 4.</b> Perhitungan Jarak Antar Data Sinyal Untuk Tahapan Klasifikasi .....	62
<b>Lampiran 5.</b> Program Python untuk Filterisasi Sinyal .....	63
<b>Lampiran 6.</b> Program Python untuk Ekstraksi Fitur Sinyal EEG .....	66
<b>Lampiran 7.</b> Program Python untuk Pembagian Data <i>Training - Testing</i> dan Pemilihan K-Neighbors Terbaik.....	68
<b>Lampiran 8.</b> Program Python untuk Klasifikasi KNN dengan Nilai K terbaik .....	70

*(Halaman ini sengaja dikosongkan)*

## DAFTAR NOTASI

### **IIR *Butterworth***

$x_n$	<i>Input signal</i>
$y_n$	<i>Output signal</i>
$a_l$	Koefisien <i>feed forward</i>
$b_k$	Koefisien <i>feed back</i>
$N$	Orde <i>feed forward</i>
$M$	Orde <i>feed back</i>
$\omega$	Nilai <i>magnitude</i>

### **Ekstraksi Fitur**

E	Energi signal
V	<i>Variance</i> Signal
$X_n$	Nilai Sampel ke-n
N	Banyaknya Sampel pada signal
En	<i>Entropy</i> Signal

*(Halaman ini sengaja dikosongkan)*



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Otak manusia merupakan struktur pusat pengaturan yang memiliki volume sekitar 1.350cc dan terdiri atas 100 juta sel saraf atau neuron (Beatty, 2001). Otak manusia dapat melakukan berbagai aktivitas baik itu normal maupun abnormal. Kondisi abnormal dapat terjadi pada gangguan neurologis dan ketidakseimbangan akibat pengaruh obat-obatan, seperti kejang pada epilepsi dan demensia (Teplan, 2002). Jutaan sel yang terdapat pada otak memiliki proses seperti aliran listrik, yang mampu mengantarkan jutaan informasi perdetik. Lepasnya muatan listrik pada sel-sel otak secara tiba-tiba dengan jumlah yang berlebihan dan tidak teratur mengakibatkan penerimaan dan pengiriman rangsang antara bagian-bagian otak dan dari otak ke bagian-bagian tubuh lain jadi terganggu. Hal inilah menyebabkan terjadinya epilepsi.

Epilepsi merupakan suatu manifestasi gangguan fungsi otak dengan berbagai etiologi, dengan gejala tunggal yang khas, yaitu kejang berulang lebih dari 24 jam yang diakibatkan oleh lepasnya muatan listrik neuron otak secara berlebihan dan paroksismal serta tanpa provokasi (Engel & Pedley, 2008). Tiap Individu yang mengalami epilepsi mempunyai risiko yang bermakna untuk mengalami kekambuhan kejang. Waktu munculnya kejang terjadi secara mendadak, tidak disertai demam dan tidak dapat diprediksi. Pada penderita epilepsi, kejang dapat terjadi pada sebagian tubuh maupun kejang diseluruh tubuh yang dapat menyebabkan kehilangan kesadaran, cedera fisik atau bahkan dapat menyebabkan kematian (Shoeb, & Guttag, 2010).

Data organisasi kesehatan dunia (WHO, 2018) menunjukkan bahwa epilepsi menyerang 50 juta dari penduduk dunia. Prevalensi epilepsi rata-rata mencapai 8,2 per 1000 penduduk. Sementara kasus baru epilepsi lebih banyak terjadi di negara berkembang termasuk Indonesia yang rata-rata diperkirakan mencapai 114 per

100.000 penduduk per tahun. Pada negara maju kasus baru terjadi rata-rata 50 per 100.000 penduduk per tahun. Prevalensi epilepsi cukup tinggi pada bayi dan anak-anak, menurun pada dewasa muda dan pertengahan, kemudian meningkat lagi pada kelompok usia lanjut (Perhimpunan Dokter Spesialis Saraf Indonesia (Kusumastuti, Gunadharma, & Kustiowati, 2011).

Kejang adalah kejadian epilepsi dan merupakan ciri epilepsi yang harus ada, tetapi tidak semua kejang merupakan manifestasi epilepsi. Selain karna epilepsi, kejang juga dapat diakibatkan oleh penyebab lain termasuk reaksi obat, kondisi jantung, stres psikis, atau bahkan kurang tidur yang ekstrim. Diagnosa pasti epilepsi adalah dengan menyaksikan secara langsung terjadinya serangan, namun serangan epilepsi jarang bisa disaksikan langsung oleh dokter, sehingga diagnosa epilepsi hampir selalu dibuat berdasarkan keterangan dari orang sekitar penderita yang biasanya sering kali tidak khas dan tidak akurat, sedangkan penderitanya sendiri tidak tahu sama sekali bahwa ia baru saja mendapat serangan epilepsi. Salah satu metode yang membantu mendiagnosa kejang epilepsi adalah secara obyektif dengan menggunakan peralatan *electroencephalogram* (EEG).

EEG merupakan metode atau teknik perekaman aktivitas listrik pada otak secara spontan dalam periode waktu yang singkat, biasanya berkisar antara 20 menit sampai 40 menit. Hasil dari perekaman tersebut berupa sinyal gelombang otak. Sinyal tersebut didapatkan dari beberapa elektroda yang dipasang pada kulit kepala. Sinyal EEG yang diukur akan memiliki amplitudo sekitar 10  $\mu$ V sampai 100  $\mu$ V dan frekuensi dikisaran 1 Hz sampai sekitar 100 Hz (Subbha, dkk., 2010). Secara tradisional, rekaman EEG diperiksa secara visual oleh ahli neurologi yang terlatih untuk mendeteksi kejang epilepsi atau kelainan lainnya yang dialami oleh pasien. Analisis sinyal EEG secara visual untuk mendiagnosa epilepsi tidaklah cukup karena sampai saat ini tidak ada kriteria atau ciri-ciri spesifik yang dievaluasi oleh para ahli. Selain itu, kesalahan dalam mendiagnosa penyakit dapat menyebabkan kematian bagi pasien. Sinyal EEG memiliki distribusi non-gaussian,

non-stasioner dan memiliki sifat non-linier, sehingga beberapa teknik komputerisasi otomatis telah digunakan untuk meneliti hasil rekaman EEG (Kumar, & Kolekar, 2014).

Klasifikasi merupakan salah satu metode *data mining* yang dapat digunakan meminimalisir kesalahan dalam mendiagnosa epilepsi berdasarkan sinyal EEG. Penelitian mengenai diagnosa atau deteksi epilepsi telah menggunakan beberapa metode data mining. Kumar dan Kolekar (2014) melakukan penelitian mengenai deteksi kejang epilepsi dengan menggunakan metode *Discrete Wavelet Transform* (DWT) sebagai ekstraksi fitur dan metode *Support Vector Machine* (SVM) untuk klasifikasi mendapatkan rata-rata akurasi, sensitivitas, dan spesifisitas terbaik masing-masing sebesar 97,5%, 98%, dan 96%. Penelitian yang sama juga dilakukan oleh Guo, dkk. (2011) dengan menggunakan *Genetic programming based* sebagai metode ekstraksi fitur dan metode klasifikasi dengan KNN didapatkan akurasi sebesar 93,5%. Penelitian mengenai deteksi epilepsi juga pernah dilakukan oleh Napitupulu, dkk (2017) dengan menggunakan *Weighted Permutation Entropy* (WPE) sebagai metode ekstraksi fitur dan metode *Support Vector Machine* (SVM) untuk klasifikasi mendapatkan rata-rata akurasi sebesar 91,88%.

Pada penelitian ini akan dijelaskan mengenai diagnosa kejang epilepsi menggunakan metode IIR *butterworth* Filter dan KNN. IIR merupakan salah satu metode filter yang menghasilkan panjang gelombang yang sama dengan sinyal aslinya. EEG terdiri atas lima bagian frekuensi yaitu *delta*, *theta*, *alpha*, *beta*, dan *gamma*. Dari kelima *sub-band* tersebut hanya *theta*, *alpha*, dan *beta* yang membawa informasi kejang (Kumar, & Kolekar, 2014). Sehingga pada tugas akhir ini metode IIR *butterworth* filter digunakan untuk mendapatkan sinyal dengan ketiga *sub-band* tersebut. Sedangkan KNN merupakan salah satu metode klasifikasi paling tua yang berasal dari konsep NN, paling sederhana dan populer namun memiliki kinerja yang baik yang mampu menyamai kinerja metode lain yang lebih rumit seperti PNN (Yang, dkk., 2007), (Tamouk, & Allahakbari, 2012), (Statnikov, dkk., 2005).

Alotaiby, dkk (2015) melakukan penelitian mengenai diagnosa kejang epilepsi menggunakan metode IIR *Butterworth* untuk filterisasi sinyal dan klasifikasi menggunakan SVM didapatkan nilai akurasi, spesifisitas, sensisitivitas masing-masing sebesar 93,15, 98,279, dan 87,04. Kemudian, Karlık & Hayta (2014) melakukan perbandingan klasifikasi dengan metode *Artificial Neural Networks* (ANN), *Naive Bayesian*, *k-Nearest Neighbor* (KNN), *Support Vector Machines* (SVM) dan *k-Means* menyimpulkan bahwa metode KNN lebih baik dari metode-metode lainnya. Data yang digunakan dalam penelitian ini adalah data rekaman EEG otak manusia yang diperoleh dari Klinik für Epileptologie Universität Bonn. Diharapkan nilai akurasi pada penelitian ini lebih baik dari penelitian sebelumnya sehingga dapat digunakan dalam mendeteksi kejang epilepsi.

## 1.2 Rumusan Masalah

Kejang merupakan salah satu ciri khas epilepsi, tetapi tidak semua kejang merupakan manifestasi dari epilepsi. EEG merupakan metode atau teknik perekaman aktivitas listrik pada otak dengan hasil rekaman berupa sinyal. Sinyal EEG memiliki distribusi non-gaussian, non-stasioner, dan non linear, sehingga metode visual tidaklah cukup untuk mendiagnosa kejang epilepsi berdasarkan sinyal EEG. Selain itu sinyal EEG terdiri atas lima *sub-band* frekuensi yaitu frekuensi *delta*, *theta*, *alpha*, *beta*, dan *gamma* namun hanya *sub-band* , *theta*, *alpha*, dan *beta* yang membawa informasi kejang. Sehingga dibutuhkan metode komputerisasi dalam mendiagnosa kejang epilepsi. Berdasarkan uraian tersebut permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana hasil filterisasi dan ekstraksi variabel dari sinyal EEG dan bagaimana hasil klasifikasi sinyal EEG dengan menggunakan KNN.

## 1.3 Tujuan

Tujuan yang dapat diambil dari penelitian ini adalah sebagai berikut.

1. Mendapatkan data sinyal hasil filterisasi dengan gelombang *theta*, *alpha*, dan *beta* dan ekstraksi fitur dari masing-masing gelombang.
2. Mengetahui hasil klasifikasi kejang epilepsi berdasarkan fitur EEG yang telah diekstrak.

#### **1.4 Manfaat**

Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut.

1. Membantu ahli neurologi dalam mengambil keputusan mengenai diagnosa penyakit kejang epilepsi berdasarkan sinyal EEG sehingga dapat segera dilakukan tindakan preventif untuk pengobatan pasien.
2. Menambah referensi dalam penggunaan metode IIR *Butterworth* Filter dalam filterisasi sinyal EEG serta klasifikasi kejang epilepsi menggunakan metode KNN.

*(Halaman ini sengaja dikosongkan)*

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 *Infinite Impulse Response (IIR) Filter*

Filter atau tapis adalah suatu sistem yang berfungsi untuk menyaring sinyal, sebagian sinyal akan dibiarkan lewat, sebagian yang lain akan ditahan. Berdasarkan sinyal yang diproses, filter dibagi menjadi dua, yaitu filter analog dan filter digital. Secara umum, filter digital adalah sama dengan filter analog, hanya saja sinyal input dan sinyal *output*-nya adalah sinyal digital. Filter digital memiliki banyak kelebihan dibandingkan dengan pasangannya filter analog, baik dalam performa yang lebih tinggi dengan *transition zone* yang lebih kecil, ketahanan, serta fleksibilitas dalam menentukan *range* kerjanya (Smith, 1997). Karena faktor-faktor ini, filter digital merupakan elemen penting dalam bidang DSP (*Digital Signal Processing*).

Terdapat dua metode untuk mendesain sebuah filter digital. Metode pertama dengan menggunakan proses konvolusi antara sinyal input dengan *impulse response* dari filter yang dikehendaki, filter jenis ini disebut filter FIR (*Finite Impulse Response*). Metoda kedua adalah dengan proses rekursif, yang merupakan kelanjutan dari metoda konvolusi. Bila dalam proses konvolusi perhitungan dilakukan dengan hanya menggunakan sampel input saja, maka dalam proses rekursif perhitungan dilakukan dengan sampel input yang dijumlahkan dengan sampel *output* sebelumnya. Hal ini membuat metode rekursif menjadi sangat panjang dan mendekati titik tak berhingga (*infinity*), oleh karena itu filter jenis ini disebut filter IIR (*Infinite Impulse Response*).

Pengertian sederhana untuk *IIR* filter disini adalah bahwa *output* filter merupakan fungsi dari kondisi input sekarang. Konsep ini kemudian lebih dikenal sebagai *recursive filter*, yang mana

melibatkan proses *feed back* dan *feed forward*. Dalam bentuk persamaan beda yang menghubungkan *input* dengan *output* dinyatakan dengan persamaaan (2.1) berikut ini.

$$y_m = \sum_{l=0}^N a_l y_{m-l} + \sum_{k=0}^M b_k x_{m-k} \quad (2.1)$$

dengan,

$a_l$  = koefisien *feed forward* ke- $l$

$b_k$  = koefisien *feed back* ke- $k$

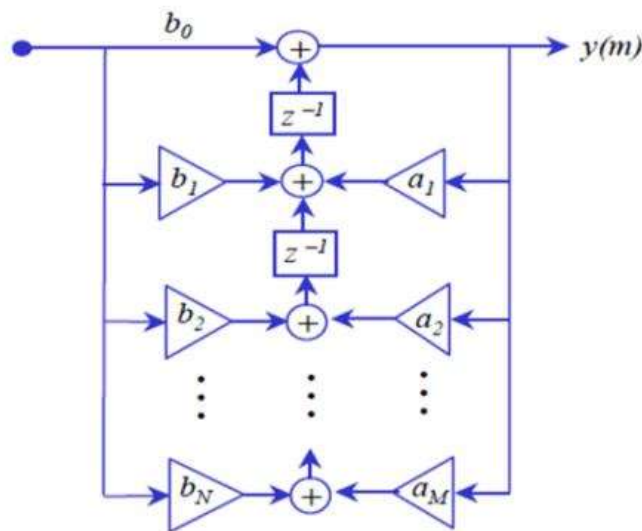
$N$  = orde *feed forward*

$M$  = Orde *feed back*

$x_m$  = *input* signal

$y_m$  = *output* signal.

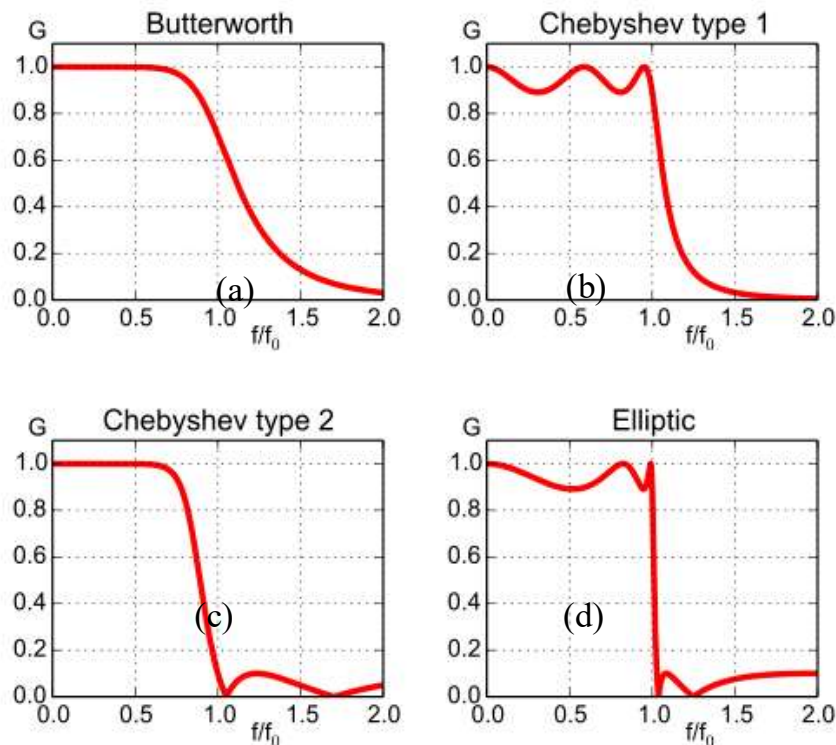
Untuk merealisasikan ke dalam sebuah program simulasi atau perangkat keras maka bentuk persamaan diatas dapat disederhanakan ke dalam diagram blok seperti pada Gambar 2.1.



**Gambar 2.1** Diagram blok *Filter IIR*  
(Sumber: Podder, dkk., 2014)

Konfigurasi *Butterworth* adalah salah satu konfigurasi standar dari filter rekursif baik dalam bentuk analog maupun digital.





**Gambar 2.2** Tipe *Magnitude Respon* dari Konfigurasi *Filter butterworth* (a), *Chebyshev* (b dan c), dan *Elliptical* (d)  
(Sumber: Podder, dkk., 2014)

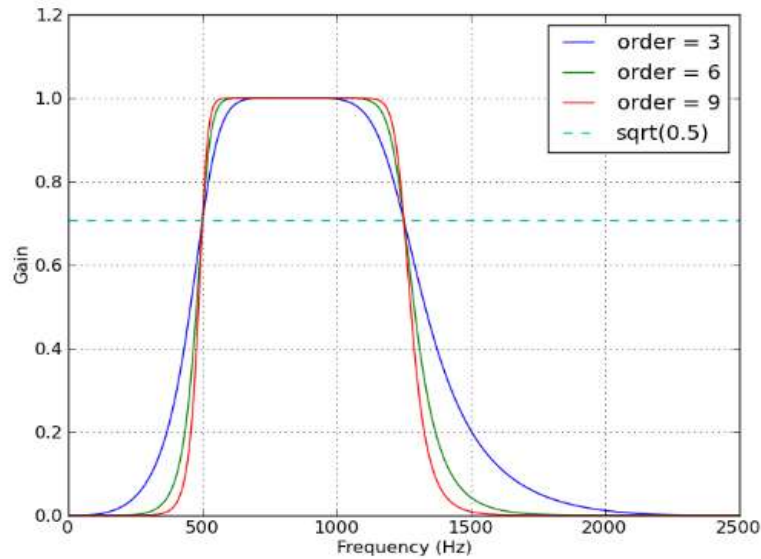
Konfigurasi ini menekankan pada aproksimasi karakteristik *lowpass* dengan hasil *respons* yang mendekati titik nol dengan halus dan rata (*smooth* and *flat*) (Soliman, 1990). Filter *Butterworth* didefinisikan melalui persamaan *magnitude function*  $H(\omega)$  sebagai berikut:

$$H(\omega) = \frac{1}{\sqrt{1+\omega^{2N}}} \quad (2.2)$$

dengan  $N$  adalah nilai orde filter.

Jelas dari rumus di atas bahwa *magnitude function Butterworth* adalah fungsi frekuensi ( $\omega$ ) yang menurun secara monoton, dengan nilai maksimumnya dari *unity* terjadi pada saat  $\omega = 0$ . Untuk  $\omega = 1$ , nilai *magnitude* adalah sama dengan

$1/\sqrt{2}$  untuk semua nilai  $N$ . Dengan demikian, filter *Butterworth* dalam bentuk normal memiliki frekuensi *cutoff* sebesar 3 dB. Gambar 2.3 menunjukkan plot dari karakteristik *magnitude* dari filter *Butterworth* untuk beberapa tingkatan orde.



**Gambar 2.3** Tipe *Magnitude Respon* dari Konfigurasi *Filter butterworth* untuk Berbagai Tingkat Orde

## 2.2 Ekstraksi Fitur

Ekstraksi fitur merupakan tahapan yang digunakan untuk menghasilkan fitur-fitur yang akan digunakan sebagai variabel prediktor dalam metode klasifikasi. Berikut merupakan beberapa fitur yang dapat diekstrak dari sinyal EEG (Hamad, dkk., 2016) dan (Kumar, & Kolekar, 2014).

### 1. *Energy*

Energi dari sinyal didefinisikan sebagai jumlah kuadrat dari nilai modulus sampel. *Energy* dari sinyal dapat dirumuskan sebagai berikut.

$$E = \sum_{n=1}^N |X_n|^2 \quad (2.3)$$

Dimana  $X_n$  adalah nilai sampel, dan N adalah jumlah sampel masing-masing *sub-band*.

## 2. *Variance*

Pengukuran *Variance* menunjukkan seberapa jauh persebaran dari sekumpulan data. *Variance* dari sekumpulan data dapat dihitung dengan menggunakan Persamaan (2.4).

$$V = \frac{\sum_{n=1}^N (X_n - \bar{X})^2}{N-1} \quad (2.4)$$

Dengan  $\bar{X}$  adalah rata-rata dari data sampel X, dan N adalah jumlah sampel.

## 3. Maksimum

Maksimum adalah nilai sampel terbesar dari setiap *sub-band*

## 4. *Minimum*

Mimumum adalah nilai sampel terkecil dari setiap *sub-band*

## 5. *Entropy*

*Entropy* merupakan pengukuran ketidakpastian secara numerik dari *outcome* letak sinyal yang mengandung ribuan bit informasi. Secara matematis *Entropy* dirumuskan dengan Persamaan (2.5).

$$(EN) = \sum_{n=1}^N X_n^2 \log(X_n^2) \quad (2.5)$$

## 2.3 *K-Nearest Neighbor (KNN)*

*Nearest Neighbor* (NN) adalah salah satu metode yang banyak digunakan dalam *Data Mining*. Metode NN diklasifikasikan sebagai *lazy laearner*, karena metode ini menunda proses pelatihan atau tidak melakukan sama sekali, sampai ada data uji yang ingin diketahui label kelasnya. Ketika ada data uji yang ingin diketahui label kelasnya, maka metode NN baru akan menjalankan algoritmanya (Han, Kamber, & Pei, 2012). *K-Nearest Neighbors* merupakan salah metode satu paling tua yang berasal dari konsep NN, paling sederhana dan populer namun memiliki

kinerja yang baik yang mampu menyamai kinerja metode lain yang lebih rumit seperti PNN (Yang, dkk., 2007), (Tamouk & Allahakbari, 2012), (Statnikov, dkk., 2005). Algoritma KNN dijalankan berdasarkan konsep kedekatan atau jarak data dengan data lainnya, dengan nilai K menunjukkan jumlah data terdekat yang digunakan untuk pengklasifikasian kelas data.

Pemilihan nilai K merupakan hal yang sangat mempengaruhi kinerja metode KNN (Prasetyo, 2014). Nilai K yang terlalu kecil atau terlalu besar akan mengakibatkan hasil prediksi sensitif terhadap gangguan/*noise* atau adanya risiko bahwa data yang dipilih tidak relevan karena mayoritas data yang dipilih bisa dari data yang berlainan kelas atau terlalu jauh dari data yang ingin dikenali label kelasnya. Isu lain yang berkaitan dengan pemilihan nilai K adalah pemilihan nilai K dengan nilai ganjil atau genap. Jika nilai K ganjil, risiko terdapat dua kelas atau lebih memperoleh jumlah suara sama sangat kecil. Sedangkan ketika nilai K genap juga ada risiko dua kelas atau lebih memiliki jumlah suara yang sama. KNN biasanya diaplikasikan dengan validasi 10-lipatan (*10-fold validation*), menggunakan aturan jarak terdekat (*tie-break rules*) untuk mengatasi masalah pemilihan kelas karena jumlah suara yang sama kuat, dan pemberian bobot pengaruh berdasarkan jarak untuk memberikan pengaruh yang lebih kecil terhadap data yang terletak jauh dari data uji.

Metode KNN tidak membutuhkan pembangunan model dikarenakan model yang akan diujikan yaitu keseluruhan data latih. Dalam melakukan prediksi terhadap data uji maka data latih mulai digunakan untuk mencari kemiripan data sesuai jumlah k yang ditentukan terlebih dahulu. Algoritma KNN mempunyai kelebihan yaitu mudah dimengerti dan diterapkan, proses pelatihan berlangsung sangat cepat, tangguh terhadap data pelatihan yang terganggu (*noise*) serta bisa diimplementasikan pada aplikasi

dengan banyak jumlah kelas label (Bhavsar & Ganatra, 2014). Sedangkan kelemahan yang dimiliki knn yaitu sensitif terhadap struktur data lokal (Duda, & Hart, 1973), terbatasnya memori penyimpanan, serta proses klasifikasi sangat lambat (Bhavsar & Ganatra, 2014).

Algoritma *K-Nearest Neighbor* mempunyai langkah langkah dalam klasifikasi sebagai berikut.

1. Menentukan nilai K  
Penentuan nilai K dimulai dari  $K = 1$  saat menggunakan data uji untuk tingkat kesalahan pengklasifikasian. Nilai K dapat ditambah sampai mendapatkan nilai kesalahan yang paling minimum (Han, Kamber, & Pei, 2012).
2. Menghitung jarak antara data latih dengan data data uji  
Jika data merupakan data numerik salah metode yang dapat digunakan dalam menghitung jarak antar data latih dengan data uji adalah dengan menggunakan jarak *euclidean*, dengan rumus sebagai berikut (Han, Kamber, & Pei, 2012).

$$dist(x_1, x_2) = \sqrt{\sum_{i=1}^N (x_{1i} - x_{2i})^2} \quad (2.6)$$

Sedangkan jika data kategorikal, dilakukan dengan cara membandingkan nilai antara atribut a dan b. jika kedua atribut identik maka akan diberi nilai 0, dan jika tidak identik akan diberi nilai 1 (Han, Kamber, & Pei, 2012). Namun sebelumnya nilai dari setiap atribut harus dinormalisasi terlebih dahulu menggunakan metode *Min-max normalization* dengan mengubah nilai  $v$  dari atribut numerik  $A$  menjadi  $v'$  pada range  $[0,1]$  dengan mengkomposisikan rumus seperti Persamaan (2.7).

$$v' = \frac{v - \min_A}{\max_A - \min_A} \quad (2.7)$$

Dimana  $\min_A$  dan  $\max_A$  adalah nilai paling kecil dan paling besar dari atribut A.

### 3. Mengurutkan jarak terdekat

Data dari semua perhitungan nilai jarak akan diurutkan bersarkan nilai yang terkecil untuk masing-masing nilai K. sehingga akan diperoleh jarak yang paling dekat antara data latih dengan data uji.

## 2.4 *Stratified K-Fold Cross-Validation*

Pada *stratified k-fold cross-validation* data akan dipartisi secara acak menjadi  $k$  bagian atau *folds* dengan jumlah sampel sama yaitu  $D_1, D_2, \dots, D_k$  dan masing masing-masing *fold* memiliki proporsi yang sama antar kelas data. Validasi menggunakan *training* dan *testing* dilakukan sebanyak  $k$  kali. Pada iterasi ke- $i$ , partisis  $D_i$  akan diatur sebagai data *testing* dan partisi lainnya akan digunakan sebagai data *training*. Artinya, pada iterasi yang pertama, partisi  $D_2, D_3, \dots, D_k$  akan menjadi data *training* untuk mendapatkan model pertama yang akan diuji dengan partisi  $D_1$ . Pada iterasi kedua partisi  $D_1, D_3, \dots, D_k$  akan menjadi data *training* kemudian  $D_2$  akan menjadi data *testing*, begitu seterusnya sampai dengan  $D_k$  yang akan menjadi data *testing* (Han, Kamber, & Pei, 2012).

## 2.5 Ukuran Kinerja Klasifikasi

Ketepatan prediksi dari suatu pengklasifikasian dapat dibuat dalam suatu tabel yang disebut dengan *confusion matrix* seperti pada Tabel 2.1. *Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan kinerja performansi klasifikasi pada konsep data mining. Kinerja Pengukuran yang sering digunakan untuk menghitung ketepatan

klasifikasi adalah akurasi, spesifisitas, dan sensitivitas (Hotho, Nurnberger, & Paass, 2005).

**Tabel 2.1** *Confusion Matrix*

	<b>Prediksi (+)</b>	<b>Prediksi (-)</b>	
<b>Aktual (+)</b>	TP ( <i>True Positives</i> )	FN ( <i>False Negatives</i> )	Marginal
<b>Aktual (-)</b>	FP ( <i>False Positives</i> )	TN ( <i>True Negatives</i> )	
	Marginal		Total

Akurasi merupakan persentase dokumen yang teridentifikasi secara tepat dari total dokumen dalam proses klasifikasi. Akurasi digunakan untuk menghitung ketepatan klasifikasi sebuah dokumen yang mempunyai data yang *balanced* pada tiap kategorinya. Berikut merupakan rumus dalam menghitung akurasi, spesifisitas, dan sensitivitas.

$$Akurasi = \frac{TN+TP}{TN+TP+FN+FP} \quad (2.8)$$

$$Sensitivitas = \frac{TP}{TP+FN} \quad (2.9)$$

$$Spesifisitas = \frac{TN}{TN+FP} \quad (2.10)$$

Sedangkan untuk data *imbalanced*, pengukuran ketepatan klasifikasi yang digunakan adalah *G-Mean*. *G-Mean* atau *geometric mean* merupakan rata-rata geometrik nilai *recall* dari data yang memiliki dua kategori (Sun, Kamel, & Wang, 2006). Dalam mengukur nilai performansi klasifikasi, *G-Mean* memiliki kelebihan yaitu nilai klasifikasi yang dihasilkan *robust*. Berikut merupakan rumus untuk mendapatkan nilai *G-Mean*. Selain *G-Mean* juga digunakan nilai Area Under Curve (AUC). AUC merupakan indikator performansi kurva ROC (*Receiver Operating*

*Characteristic*) yang dapat meringkas kinerja sebuah *classifier* menjadi satu nilai (Bekkar, Djemaa, & Alitouch, 2013).

$$G - mean = \sqrt{Sensitivitas \times Spesitifitas} \quad (2.11)$$

$$AUC = \frac{1}{2} (Sensitivitas + Spesitifitas) \quad (2.12)$$

## 2.6 Epilepsi

Epilepsi merupakan salah satu kelainan neurologis yang paling sering terjadi. Data organisasi kesehatan dunia (WHO) menunjukkan bahwa epilepsi menyerang 50 juta dari penduduk dunia (Brodie, 2012). Prevalensi epilepsi rata-rata mencapai 8,2 per 1000 penduduk. Sementara kasus baru epilepsi lebih banyak terjadi di negara berkembang termasuk Indonesia yang rata-rata diperkirakan mencapai 114 per 100.000 penduduk per tahun. Pada negara maju kasus baru terjadi rata-rata 50 per 100.000 penduduk per tahun.

Penyakit epilepsi adalah suatu kondisi yang dapat menjadikan seseorang mengalami kejang secara tiba-tiba dan berulang. Kerusakan atau perubahan di dalam otak diketahui sebagai penyebab pada sebagian kecil kasus epilepsi. Namun pada sebagian besar kasus yang pernah terjadi, penyebab masih belum diketahui secara pasti. Di dalam otak manusia terdapat neuron atau sel-sel saraf yang merupakan bagian dari sistem saraf. Tiap sel saraf saling berkomunikasi dengan menggunakan impuls listrik. Pada kasus epilepsi, kejang terjadi ketika impuls listrik tersebut dihasilkan secara berlebihan sehingga menyebabkan perilaku atau gerakan tubuh yang tidak terkendali.

Manifestasi klinis epilepsi dapat berupa kejang umum, kejang fokal, penurunan kesadaran, kelainan tingkah laku sampai manifestasi klinik lainnya yang penyebabnya masih sulit



dimengerti (Harsono, 1999). Adanya variasi dalam manifestasi klinis ini, maka cukup sulit membedakan jenis epilepsi secara klinis. Bangkitan epilepsi tidak selalu bersifat eksitasi atau kejang melainkan dapat juga bersifat fenomena negatif berupa menghilangnya kesadaran dan tonus otot, sehingga kadang-kadang epilepsi sulit dibedakan dengan penyakit lain yang mirip gejalanya, maka diperlukan pemeriksaan penunjang lain seperti EEG (Andarini, 2007).

## **2.7 *Electroencephalography* (EEG)**

*Electroencephalography* (EEG) adalah metode atau teknik perekaman aktifitas listrik di otak. Neuron-neuron di korteks otak mengeluarkan gelombang-gelombang listrik, yang kemudian dialirkan ke mesin EEG untuk diamplifikasi sehingga terekamlah elektroencephalogram yang ukurannya cukup untuk dapat ditangkap oleh mata pembaca EEG sebagai gelombang alfa, beta, theta dan gama (Campellone, 2006). Hasil rekaman yang dihasilkan berupa sinyal gelombang otak, yang berada pada range tegangan 5-10mV (Korba, 2015).

Sinyal EEG ditangkap melalui sensor-sensor pada elektroda yang dipasang, baik secara invasive maupun non-invasif. Pemasangan secara invasive memerlukan adanya implant elektroda-elektroda pada bagian-bagian otak yang ingin diamati. Metode ini menghasilkan sinyal yang lebih bersih dari data *noise* spasial, karena sinyal yang didapat berasal dari bagian otak yang secara langsung terkait dengan penyakit yang ingin diamati. Namun, metode ini sangat jarang dilakukan karena memerlukan usaha dan biaya yang besar, serta masalah-masalah terkait jaminan keamanannya (Nguyen, 2015).

Pemasangan secara non-invasif dilakukan dengan cara memasang *channel-channel* elektroda disekeliling tempurung

kepala, sehingga memungkinkan sinyal gelombang otak tetap diperoleh. Hal tersebut merupakan kelemahan metode non-invasif dimana antar satu *channel* dengan *channel* channel lainya selalu terdapat inferensi. Namun metode ini memiliki kemudahan pengaturan implementasi, dan juga memiliki keamanan yang terjamin, sehingga metode ini lebih dianjurkan dan sering dipakai (Nguyen, 2015). Perekaman aktivitas listrik di otak dengan menggunakan EEG merupakan salah satu metode yang dapat membantu dalam diagnose penyakit epilepsi. Diagnosa terhadap rekaman EEG sering dilakukan dalam domain waktu dengan cara melihat bentuk gelombang, ketajaman gelombang, maupun kompleksitas gelombang (Korba, 2015).

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder mengenai rekaman sinyal otak manusia dengan menggunakan EEG yang didapatkan dari *website* milik “Klinik für Epileptologie Universität Bonn”. Data tersebut terdiri dari lima set rekaman EEG yaitu set Z,O,N,F, dan S dimana masing-masing set terdiri atas 100 data rekaman EEG *single-channel* dengan durasi 23.6 detik setiap *channel*-nya. Data tersebut memiliki *sampling rate* sebesar 173,6, sehingga dengan durasi rekaman selama 23,6 akan dihasilkan jumlah sampel sebanyak  $23 \times 173,6$  sampel atau sekitar 4097 sampel setiap *channel*-nya.

Subset Z dan O diambil dari sukarelawan yang sehat namun dengan kondisi yang berbeda. Data set Z direkam dengan mata tertutup sedangkan data set O direkam dengan mata tertutup. Sub set F, N, dan S merupakan rekaman EEG otak manusia yang diambil dari pasien penderita epilepsi dengan kondisi yang berbeda. Sub set F diambil dari pasien epilepsi yang sedang tidak kejang dipusat epilepsi. Sub set N diukur dibagian *hipocampus* dengan kondisi pasien merupakan penderita epilepsi namun ketika sedang tidak kejang. Sedangkan sub set S merupakan rekaman EEG yang diambil dari penderita Kejang epilepsi. Berdasarkan uraian tersebut maka dalam penelitian ini sinyal subset Z, O, F, dan N akan dikategorikan sebagai sinyal non-kejang, sedangkan sinyal subset S akan dikategorikan sebagai sinyal kejang.

#### 3.2 Struktur Data

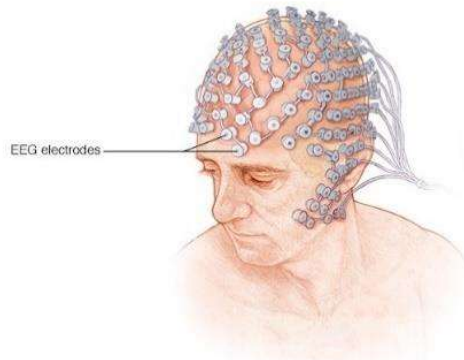
Struktur data yang akan digunakan dalam tahapan filterisasi sinyal ditunjukkan pada Tabel 3.1. Dari masing masing *Channel* pada Tabel 3.1, dilakukan Fiterisai sinyal sesuai dengan *sub-bab* (2.1), sehingga akan dihasilkan sinyal dengan frekuensi *theta* (4-8) Hz, *alpha* (8-15) Hz, dan *beta* (15-30) Hz. Pemilihan *sub-band* tersebut didasarkan pada penelitian sebelumnya yang dilakuka oleh

Kumar & Kolekar (2014) yang menyatakan bahwa dari lima *sub-band* yang terdapat pada sinyal EEG, hanya ketiga sinyal tersebut yang mengandung informasi kejang.

**Tabel 3.1** Struktur Data Sinyal Tahapan Filterisasi

Set	Amplitudo Ke-				
		1	2	...	4097
F	1	X[1,1,1]	X[1,1,2]	...	X[1,1,4097]
	2	X[1,2,1]	X[1,2,2]	...	X[1,2,4097]
	⋮				
	100	X[1,100,1]	X[1,100,2]	...	X[1,100,4097]
N	1	X[2,1,1]	X[2,1,2]	...	X[2,1,4097]
	2	X[2,2,1]	X[2,2,2]	...	X[2,2,4097]
	⋮				
	100	X[2,100,1]	X[2,100,2]	...	X[2,100,4097]
Z			⋮		
	1	X[5,1,1]	X[5,1,2]	...	X[5,1,4097]
	2	X[5,2,1]	X[5,2,2]	...	X[5,2,4097]
	⋮				
	100	X[5,100,1]	X[5,100,2]	...	X[5,100,4097]

Nilai X pada Tabel 3.1 menunjukkan nilai amplitudo sinyal. Nilai amplitudo merupakan grafik perbedaan potensial listrik di otak antara dua elektroda yang dihubungkan secara berpasangan dikulit kepala seperti yang ditunjukkan pada Gambar 3.1. Jika pada salah satu bagian otak bermuatan negatif dan satunya lagi pada bagian otak bermuatan positif, maka nilai X akan menurun dari nilai X sebelumnya. Sedangkan jika sebaliknya, Nilai X akan bertambah dari nilai sebelumnya. Jika tidak ada perbedaan potensial antara kedua elektroda maka nilai X akan cenderung stabil atau hampir sama dengan nilai sebelumnya.



**Gambar 3.1** Pemasangan Elektoda dalam Perekaman EEG  
(Sumber: Mayo Foundation For Medical Education And Research)

Selanjutnya masing-masing sinyal yang ditunjukkan pada Tabel 3.1 akan difilter menjadi 3 *sub-band* yaitu *sub-band Theta*, *Alpha*, dan *Beta*. Pada masing-masing *sub-band* dihitung beberapa fitur *wavelet* sesuai dengan sub-bab (2.2). Selanjutnya fitur tersebut akan digunakan sebagai variabel dalam klasifikasi dengan metode KNN yang ditunjukkan pada Tabel 3.2.

**Tabel 3.2** Variabel Tahapan Klasifikasi

Variabel	Keterangan	Skala Data
X1	Maksimum <i>theta</i>	Rasio
X2	Maksimum <i>alpha</i>	Rasio
X3	Maksimum <i>beta</i>	Rasio
X4	Minimum <i>theta</i>	Rasio
X5	Minimum <i>alpha</i>	Rasio
X6	Minimum <i>beta</i>	Rasio
X7	<i>Variance theta</i>	Rasio
X8	<i>Variance alpha</i>	Rasio
X9	<i>Variance beta</i>	Rasio
X10	Energi <i>theta</i>	Rasio
X11	Energi <i>alpha</i>	Rasio
X12	Energi <i>beta</i>	Rasio
X13	<i>Entropy theta</i>	Rasio
X14	<i>Entropy alpha</i>	Rasio
X15	<i>Entropy beta</i>	Rasio
Y	Kelas Sinyal EEG	Nominal
	0 = Non-Kejang Epilepsi	
	1 = Kejang Epilepsi	

Struktur data yang digunakan dalam tahapan klasifikasi ditunjukkan pada Tabel 3.3.

**Tabel 3.3** Struktur Data Tahapan Klasifikasi

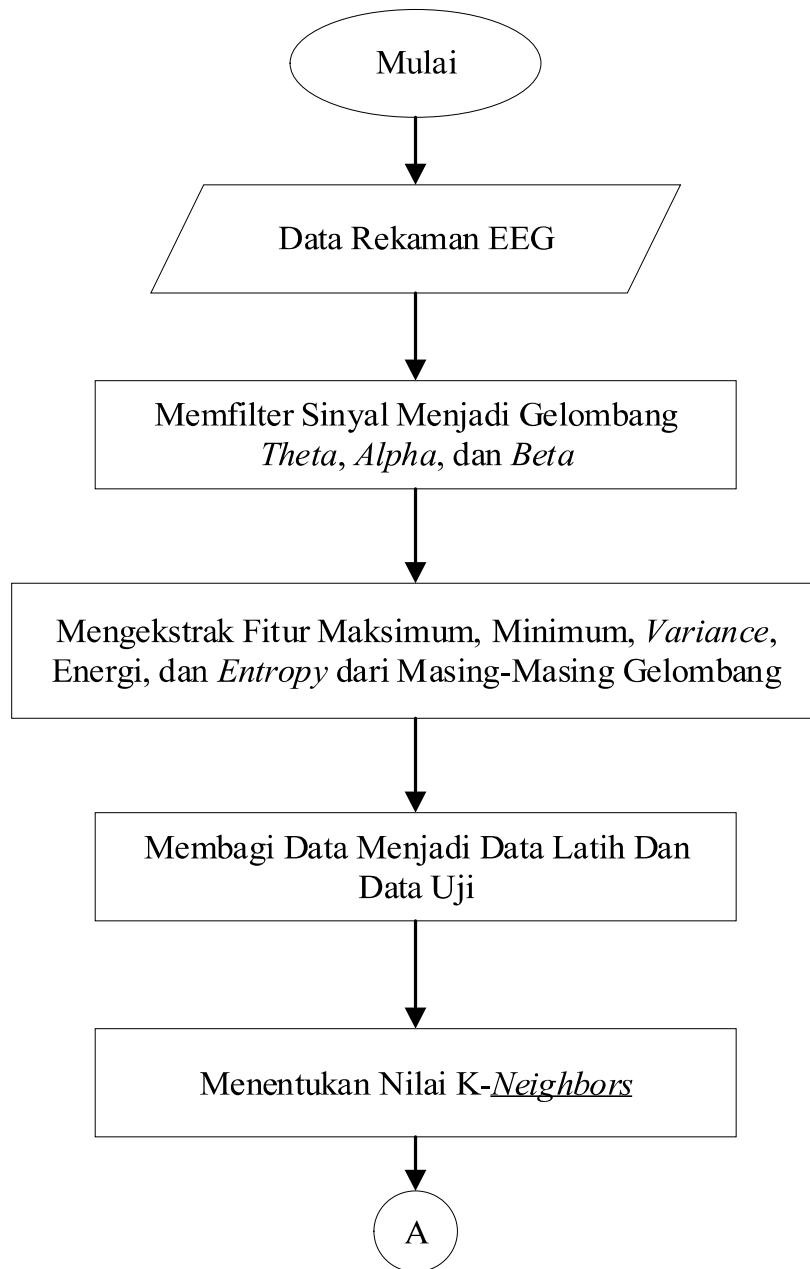
Sinyal	X1	X2	...	X15
1	X <sub>1,1</sub>	X <sub>1,2</sub>	...	X <sub>1,15</sub>
2	X <sub>2,1</sub>	X <sub>2,2</sub>	...	X <sub>2,15</sub>
3	X <sub>3,1</sub>	X <sub>3,2</sub>	...	X <sub>3,15</sub>
⋮	⋮	⋮		⋮
500	X <sub>500,1</sub>	X <sub>500,2</sub>	...	X <sub>500,15</sub>

### 3.3 Langkah Analisis

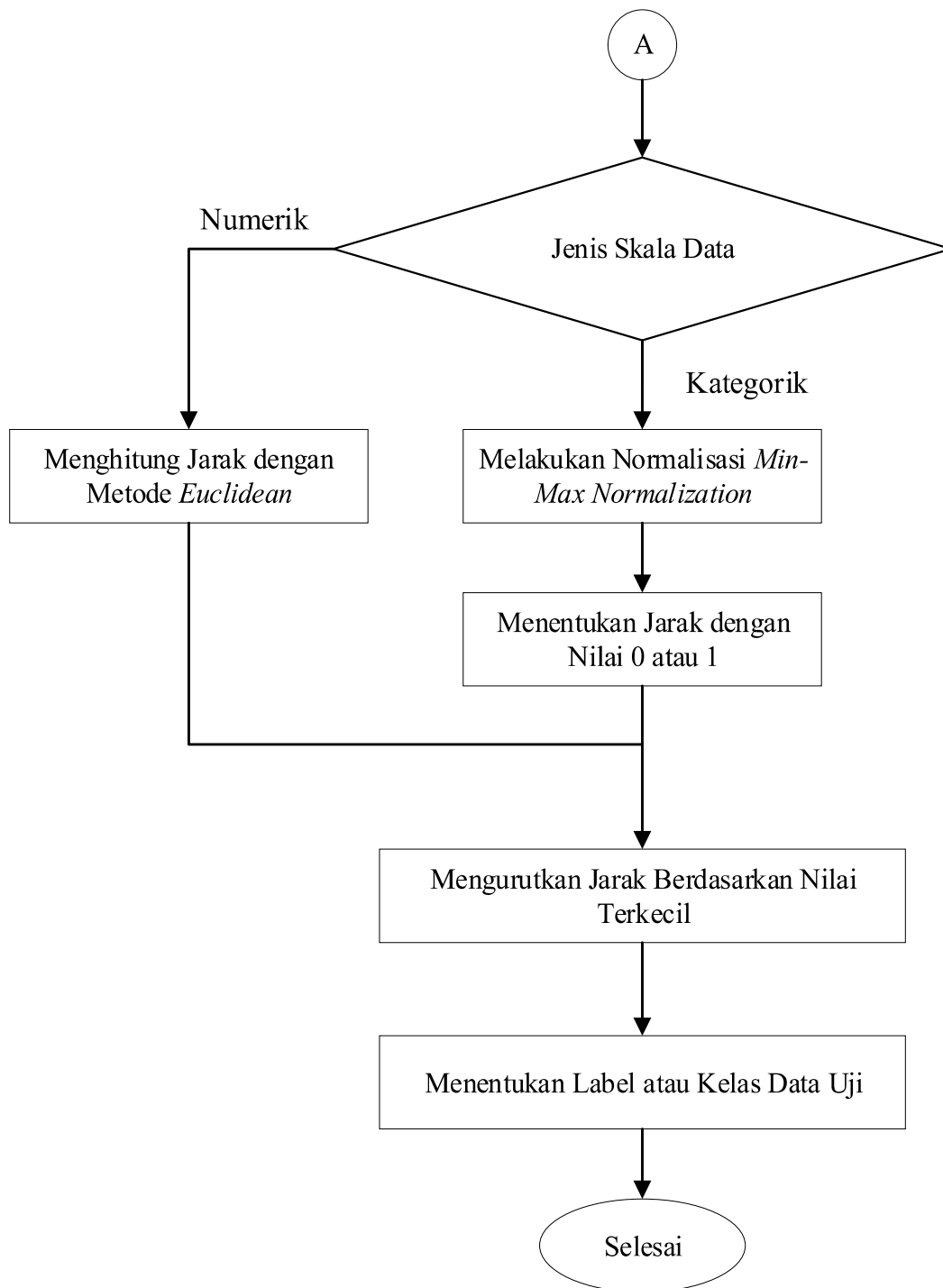
Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut.

1. Melakukan pengumpulan data sekunder rekaman EEG
2. Memfilter sinyal menjadi gelombang *theta*, *alpha*, dan *beta*
3. Mengakstrak fitur maksimum, minimum, *variance*, energi dan *entropy*
4. Melakukan klasifikasi menggunakan metode KNN
  - a. Membagi data menjadi data *training* dan data *testing* dengan menggunakan metode *stratified 5-fold cross-validation*
  - b. Menentukan nilai K
  - c. Menghitung jarak antara data *training* dengan data *testing*  
 Dalam penelitian ini penghitungan jarak dilakukan dengan menggunakan metode euclidean seperti pada Persamaan 2.6 karena semua variabel prediktor yang digunakan mempunyai skala rasio
  - d. Mengurutkan jarak berdasarkan nilai terkecil
  - e. Memilih K tetangga dengan jarak terdekat
  - f. Menentukan label data *testing*
5. Menghitung nilai ketepatan klasifikasi
6. Menarik kesimpulan dan saran

Diagram alir yang digunakan pada tugas akhir ini ditunjukkan pada Gambar 3.2.



**Gambar 3.2** Diagram Alir Penelitian



**Gambar 3.2** Diagram Alir Penelitian (Lanjutan)



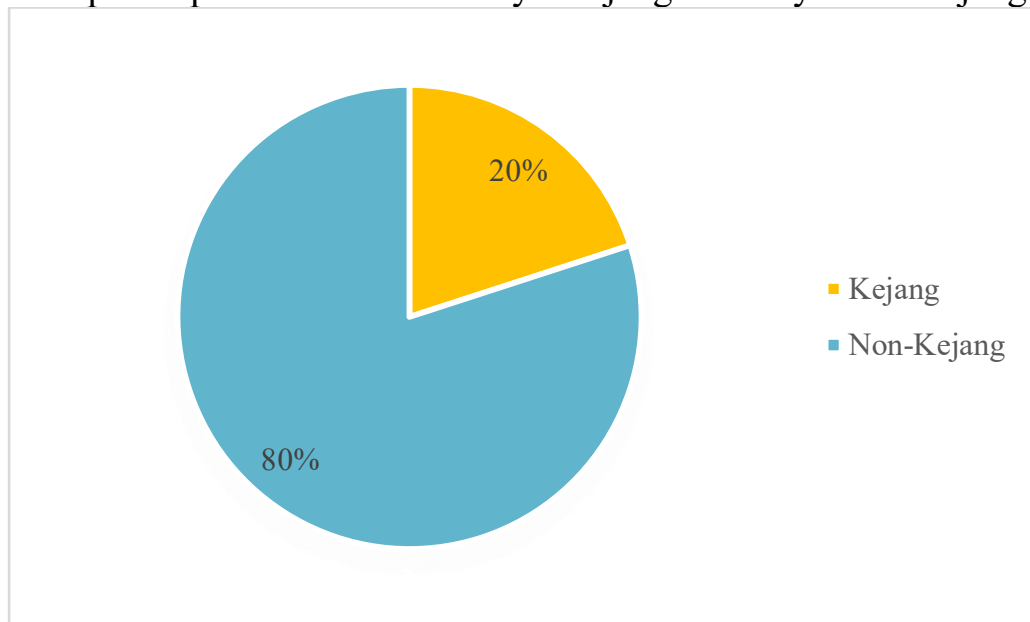
## BAB IV

### ANALISIS DAN PEMBAHASAN

Pada bab ini akan dibahas secara rinci mengenai tahapan mendiagnosa kejang epilepsi berdasarkan sinyal EEG dengan menggunakan *K-Nearest Neighbor*. Metode pembagian data untuk klasifikasi menggunakan metode *stratified 5-fold cross-validation*. Namun sebelum tahapan klasifikasi, langkah awal yang dilakukan adalah melakukan filterisasi sinyal, dan mengekstrak fitur-fitur yang akan digunakan dalam tahapan klasifikasi.

#### 4.1 Deskripsi Dataset

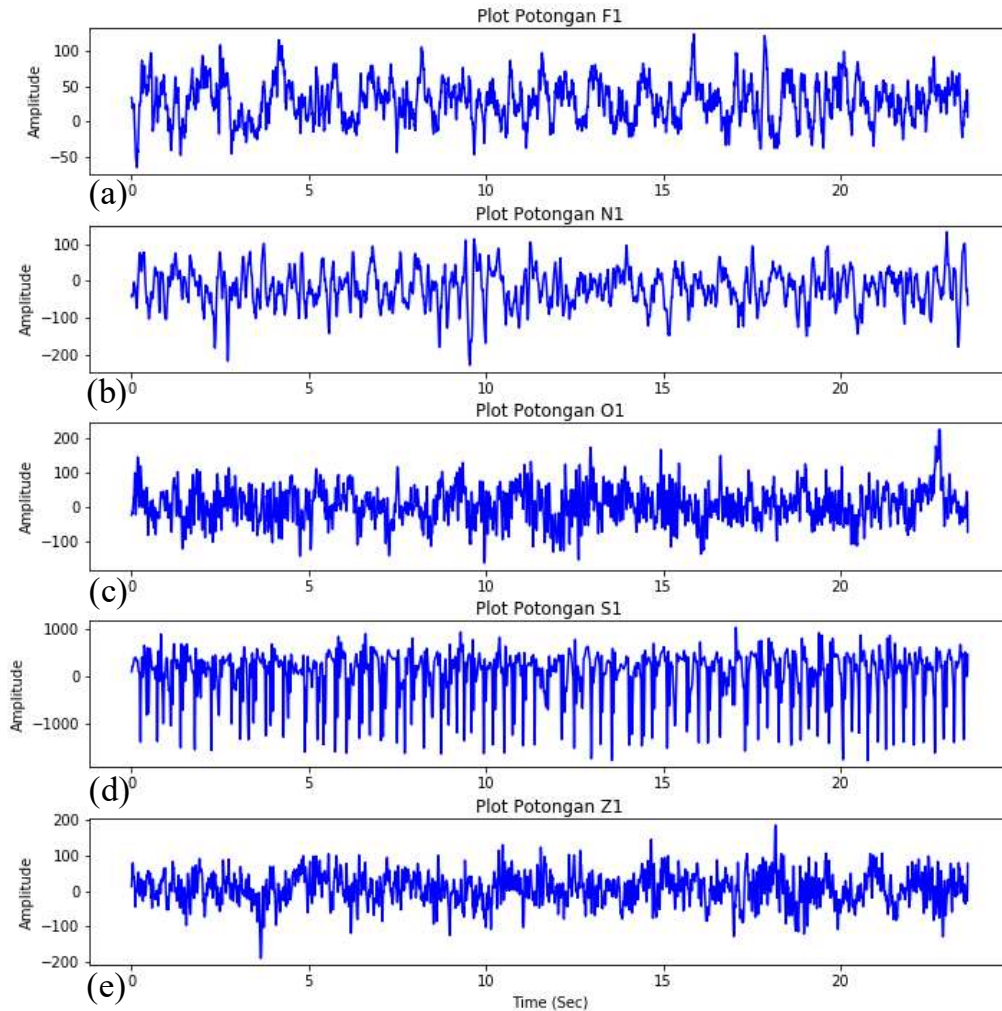
Data set yang digunakan dalam penelitian ini, terdiri dari 5 set data dengan masing masing set terdiri dari 100 potongan sinyal. Set F, N, O, dan Z dikategorikan sebagai sinyal non-kejang sementara set Z dikategorikan sebagai sinyal kejang. Berikut merupakan persentase antara sinyal kejang dan sinyal non-kejang.



**Gambar 4.1** Persentase Antara Sinyal Kejang dan Sinyal Non-Kejang

Gambar 4.1 menunjukkan bahwa sinyal kejang memiliki jumlah potongan sinyal yang lebih sedikit jika dibandingkan dengan sinyal non-kejang yaitu dengan perbandingan 20% dengan 80%.

Berdasarkan uraian tersebut maka dalam penelitian ini akan digunakan metode *stratified cross validation* dalam pembagian data *testing* dan data *training* yang akan dijelaskan pada sub bab 4.4.



**Gambar 4.2** Plot *Time Series* Data (a) Potongan F1, (b) Potongan N1, (c) Potongan O1, (d) Potongan S1, dan (e) Potongan Z1

Gambar 4.1 adalah plot *time series* dari masing-masing set. Setiap sinyal memiliki durasi perekaman selama 23,6 detik dengan *sampling rate* sebesar 173,6 yang artinya setiap detik terdapat 173.6 sampel data. Sehingga dengan durasi 23,6 detik akan dihasilkan sinyal dengan jumlah sampel sebanyak 4097. Dari

Gambar 4.2 (c) dapat dilihat bahwa data sinyal S1 memiliki *range* nilai amplitudo yang lebih besar dari kelompok data sinyal lainnya yaitu dari -1000 sampai dengan 1000. Sedangkan untuk *range* nilai amplitudo terkecil ditunjukkan oleh Gambar 4.2 (a) yang merupakan data sinyal F1. Perlu diketahui bahwa sinyal S1 merupakan data yang berasal dari kelompok kejang epilepsi, sedangkan F1 merupakan sinyal yang berasal dari kelompok penderita epilepsi tanpa kejang yang diukur di pusat epilepsi. Nilai rata-rata untuk tiap data cenderung sama yaitu berada disekitar nilai 0, namun untuk nilai varians, masing-masing kelompok data memiliki variasi yang cenderung berbeda. Dimana data sinyal S1 cenderung memiliki persebaran data yang lebih besar dari kelompok data lainnya.

Gambar 4.2 (c) dan (e) masing-masing merupakan sinyal EEG yang diambil dari pasien normal dengan keadaan mata terbuka dan mata tertutup. Kondisi dengan mata terbuka cenderung akan mengakibatkan tegangan listrik di otak lebih tinggi jika dibandingkan dengan kondisi mata tertutup. Hal tersebut sesuai dengan ditampilkan pada Gambar 4.2 (c) dan Gambar 4.2 (e) dimana sinyal dengan kondisi mata terbuka memiliki nilai amplitudo yang lebih tinggi jika dibandingkan dengan kondisi mata tertutup. Gambar 4.2 (a) dan Gambar 4.2 (b) cenderung memberikan pola persebaran data yang sama, hal ini diakibatkan karena sinyal tersebut berasal dari sinyal EEG pasien epilepsi yang direkam ketika pasien sedang tidak mengalami kejang. Selanjutnya akan dibahas mengenai filterisasi data sinyal pada sub-bab selanjutnya.

## 4.2 *Bandpass* Filter Potongan Sinyal

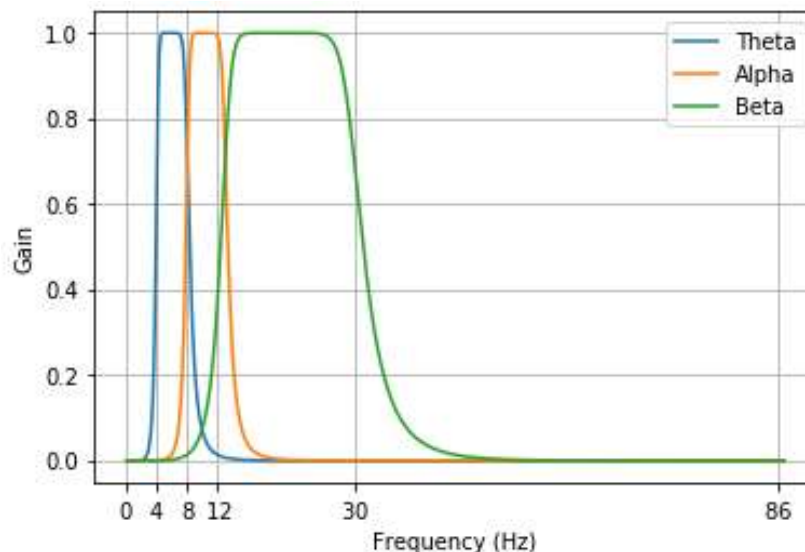
*Bandpass* filter merupakan salah satu metode yang filter yang digunakan untuk menyaring sinyal pada *range* frekuensi diatas frekuensi batas bawah (fL) dan dibawah frekuensi batas atas (fH). Pada tugas akhir ini, masing-masing data akan difilter menjadi 3 bagian yaitu sinyal dengan frekuensi *theta*, *alpha*, dan

*beta*. Filterisasi sinyal menjadi 3 bagian tersebut didasarkan pada penelitian sebelumnya yang dilakukan oleh Kumar & Kolekar (2014) yang mengatakan bahwa dari 5 kelompok sinyal yang terdapat pada sinyal EEG yaitu sinyal dengan gelombang *delta*, *theta*, *alpha*, *beta*, dan *gamma* hanya sinyal dengan gelombang *theta*, *alpha* dan *beta* yang membawa informasi kejang. Rentang gelombang setiap sinyal yang digunakan dalam penelitian ini dapat dilihat pada Tabel 4.1.

**Tabel 4.1** Rentang Frekuensi Gelombang *Tetha*, *Alpha*, dan *Beta* yang Digunakan Dalam Penelitian

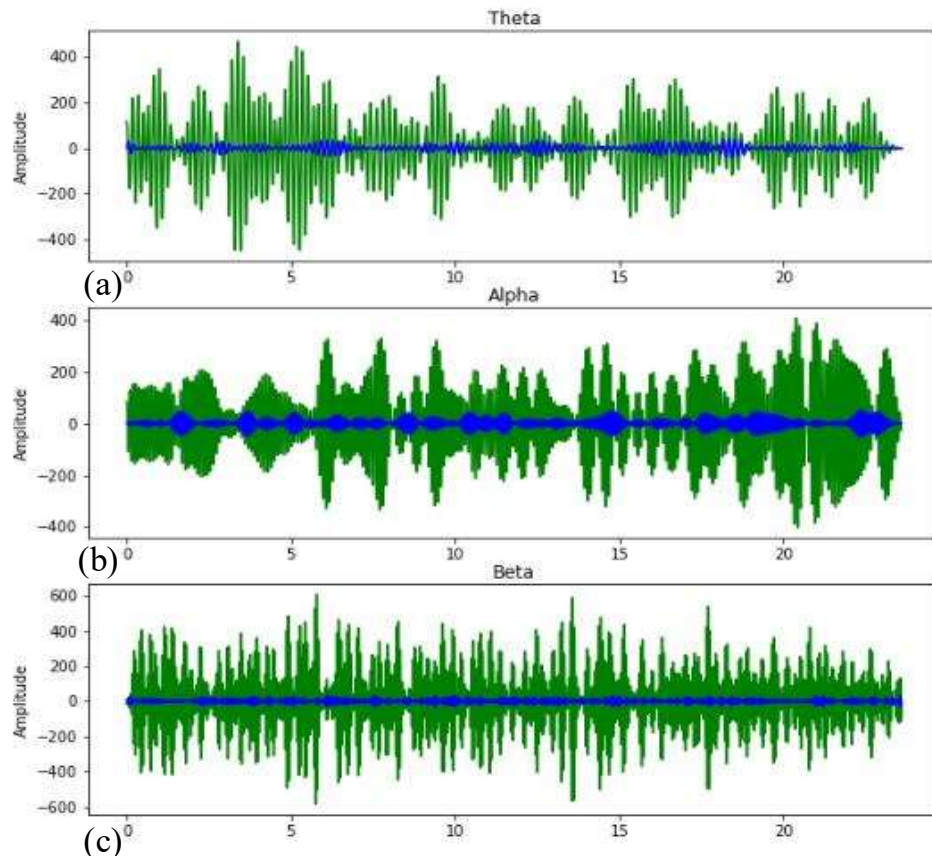
Nama gelombang	Rentang Frekuensi
<b>Theta</b>	4 - 8 Hz
<b>Alpha</b>	8 - 13 Hz
<b>Beta</b>	13 - 30 Hz

Untuk mendapatkan gelombang-gelombang tersebut, pada penelitian ini metode filter yang digunakan adalah dengan menggunakan *butterworth bandpass filter* orde 5.



**Gambar 4.3** Frekuensi Respon dari Masing-Masing Gelombang

Gambar 4.3 merupakan gambar besarnya *respon* dan frekuensi keluaran dari *band-pass* filter dengan menggunakan metode *butterworth*. Masing-masing gelombang sinyal memiliki nilai *respon* yang sama, namun untuk nilai frekuensinya, gelombang *delta* memiliki range frekuensi yang lebih besar dari kelompok gelombang lainnya. Selanjutnya hasil filterisasi sinyal kejang dan sinyal non-kejang akan ditampilkan pada Gambar 4.4.



**Gambar 4.4** Hasil Filterisasi Sinyal dengan Gelombang (a) Theta, (b) Alpha, dan (c) Beta

Gambar 4.4 merupakan plot *time Series* dari salah satu hasil filterisasi sinyal kejang yang ditunjukkan dengan garis warna hijau dan sinyal Non-kejang yang ditunjukkan dengan garis warna biru. Sinyal tersebut didapatkan dengan metode filterisasi IIR *butterworth* dengan menggunakan frekuensi batas atas dan batas

bawah pada Tabel 4.1. Dari gambar 4.3 dapat dilihat bahwa ketiga gelombang hasil filterisasi memiliki panjang dan durasi sinyal yang sama yaitu dengan jumlah sampel sebanyak 4097 dengan durasi sinyal selama 23,6 detik.

Dari Gambar 4.4 dapat dilihat bahwa sinyal hasil filterisasi masing-masing gelombang berbeda signifikan antara sinyal kejang dan sinyal non-kejang. Sinyal kejang memiliki nilai *range* Amplitudo yang lebih besar dari pada sinyal non-kejang. Selain itu, gelombang sinyal dengan gelombang *theta* memiliki kerapatan yang lebih kecil dari pada gelombang lainnya, sementara gelombang *beta* memiliki plot sinyal yang lebih rapat dari gelombang lainnya, hal ini dikarenakan *range* dari frekuensi dari masing-masing gelombang. Dimana sinyal dengan gelombang *theta* memiliki *range* frekuensi yang lebih kecil sementara gelombang *beta* memiliki *range* frekuensi yang lebih besar dari gelombang lainnya.

### 4.3 Ekstraksi Fitur Sinyal

Setelah didapatkan sinyal dengan *sub-band theta, alpha*, dan *beta*, selanjutnya adalah mengekstrak fitur dari setiap *sub-band* untuk masing-masing potongan sinyal. Fitur yang akan diekstrak terdiri fitur Maksimum, Minimum, *Variance*, Energi, dan *Entropy*. Sehingga dari ekstraksi fitur akan dihasilkan sebanyak 15 fitur atau 15 variabel prediktor yang akan digunakan untuk membedakan sinyal kejang dan sinyal non-kejang.

#### 4.3.1 Fitur Maksimum

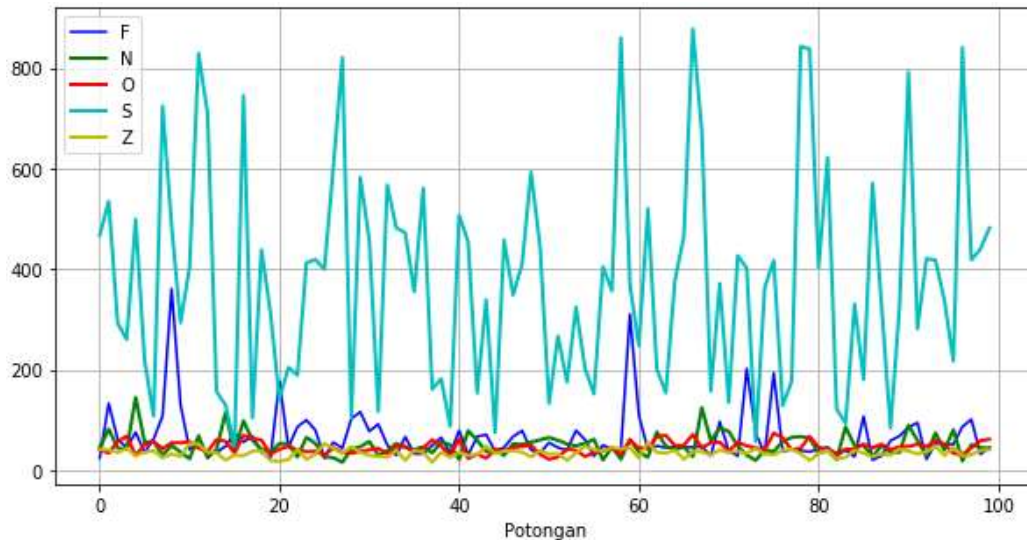
Fitur maksimum merupakan salah satu fitur yang dapat digunakan untuk membedakan antara sinyal kejang dan sinyal non-kejang. Fitur maksimum dihitung pada masing-masing *sub-band* atau kelompok gelombang. Perlu diketahui bahwa banyaknya potongan sinyal yang digunakan dalam penelitian ini adalah sebanyak 500. Sehingga fitur maksimum akan dihasilkan dataset

dengan variabel sebanyak 3 dan jumlah observasi sebanyak 500 seperti yang ditunjukkan pada Tabel 4.2.

**Tabel 4.2** Hasil Fitur Maksimum Masing-Masing *Sub-band*

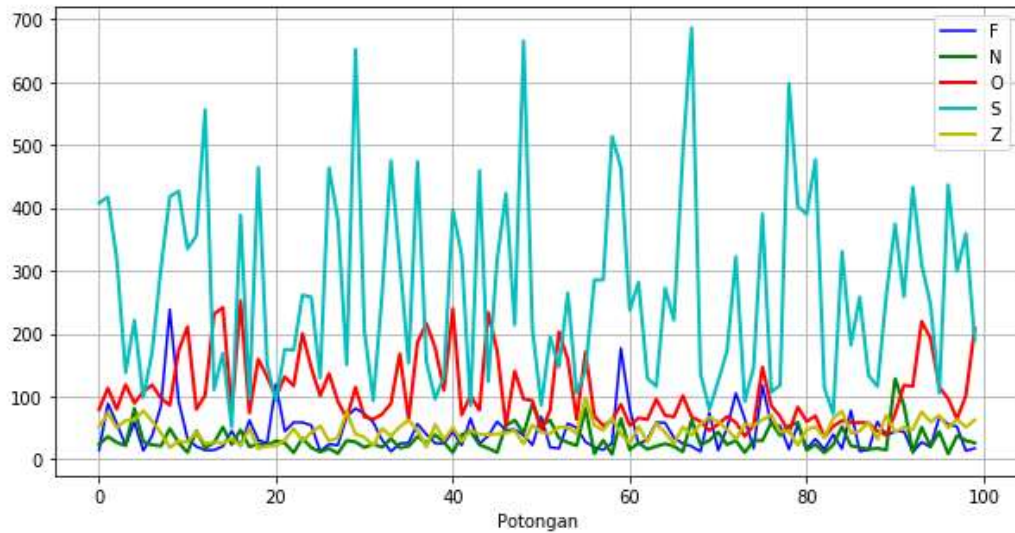
Maksimum				
Potongan	Theta	Alpha	Beta	Label
0	46,33692	29,12874	26,58152	0
1	243,7898	160,2537	124,0497	0
2	92,87788	90,81014	82,62604	0
3	80,53766	30,84423	24,09139	0
4	125,6843	96,6036	139,4821	0
⋮	⋮	⋮	⋮	⋮
498	68,10227	80,34333	80,34442	0
499	64,60916	97,48293	74,73003	0

Selanjutnya dengan menggunakan data pada Tabel 4.2 dibuat grafik untuk melihat perbedaan antara sinyal kejang dan sinyal non kejang. Gambar 4.5 merupakan grafik dari fitur maksimum untuk sub-band *Theta*, *alpha*, dan *beta*.

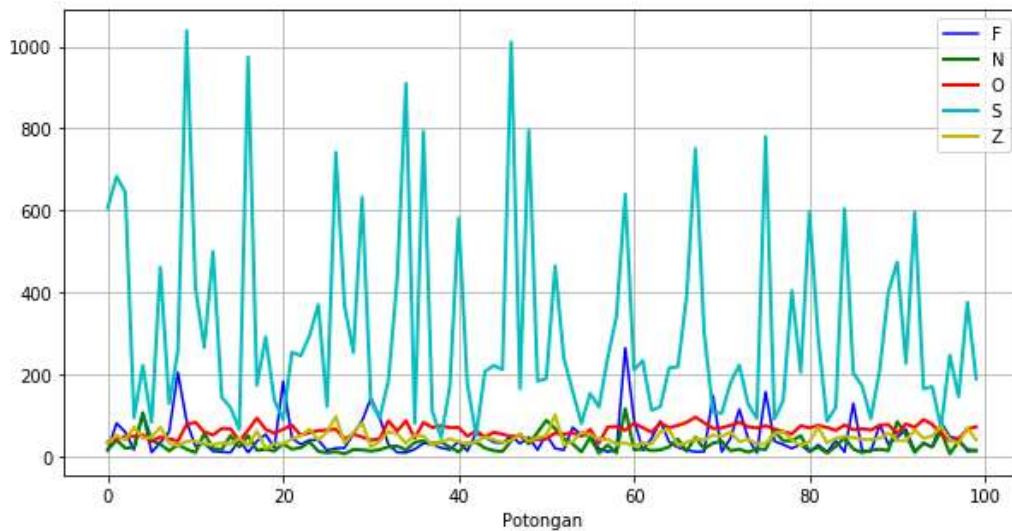


(a)





(b)



(c)

**Gambar 4.5** Plot Nilai Maksimum Gelombang *Theta* (a), *Alpha* (b), dan *Beta* (c)

Pada Gambar 4.5 dapat dilihat bahwa plot sinyal kejang (label S) cenderung berada diatas grafik grafik data sinyal non-kejang. Namun terdapat beberapa titik maksimum sinyal non-kejang dengan label F yang berada di *range* nilai maksimum label S, dan sebaliknya juga terdapat beberapa titik dari sinyal kejang yang berada di *range* nilai sinyal non-kejang. Berdasarkan hasil



ekstraksi fitur maksimum untuk ketiga gelombang, dapat dilihat bahwa gelombang *theta* dan *beta* cenderung memberikan perbedaan yang signifikan untuk antar sinyal kejang dan sinyal non-kejang. Namun untuk gelombang *alpha* terlihat bahwa plot antara sinyal kejang dan sinyal non-kejang cenderung banyak beririsan dimana nilai tersebut berasal dari sinyal dengan label 0, yaitu sinyal dari pasien normal yang diambil dengan keadaan mata terbuka.

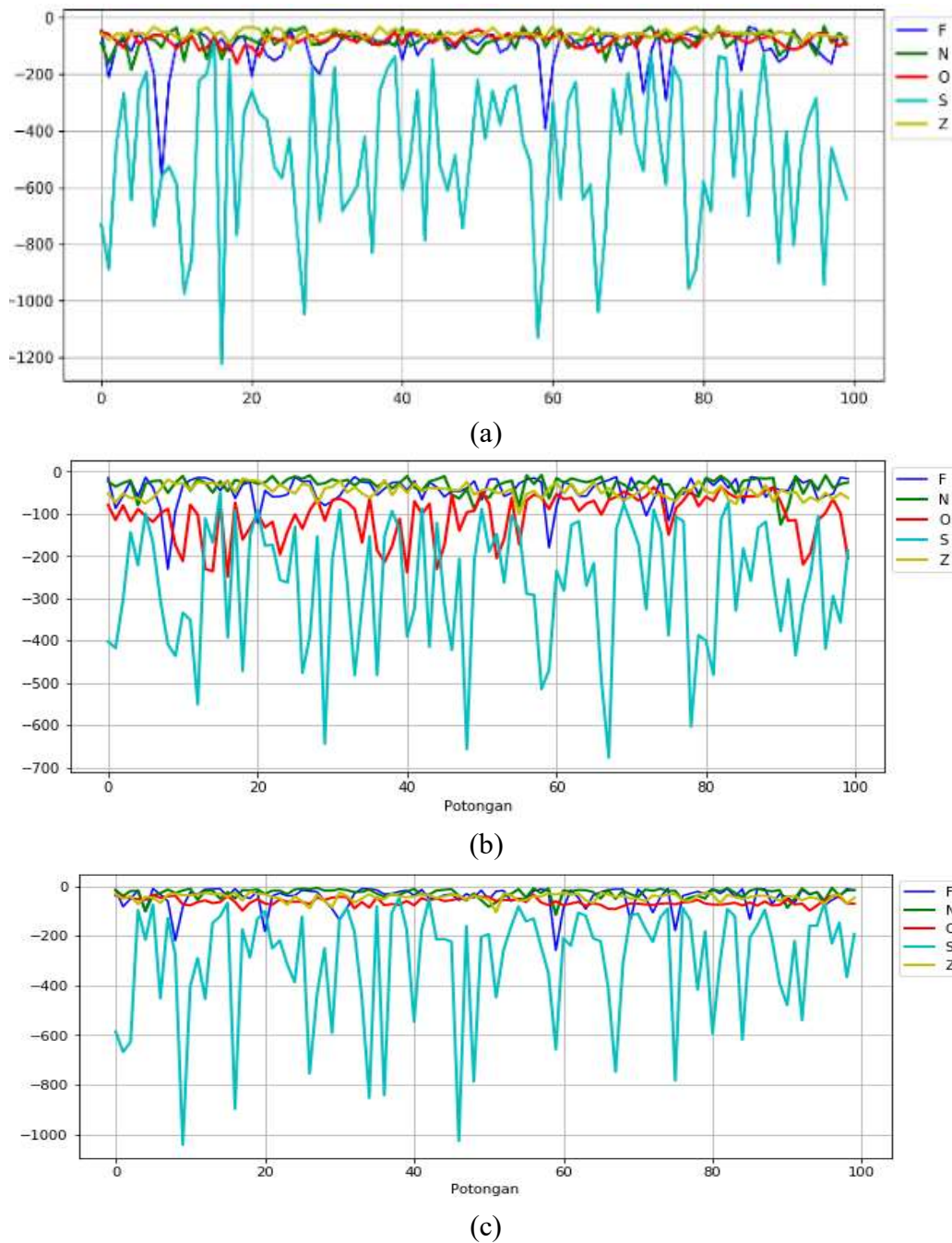
#### 4.3.2 Fitur Minimum

Selain fitur maksimum, fitur minimum juga merupakan salah satu fitur yang dapat diekstrak untuk membedakan sinyal kejang dan sinyal non-kejang. Fitur minimum merupakan nilai sampel terkecil dari setiap *sub-band*. Hasil dari perhitungan nilai minimum akan ditampilkan pada Tabel 4.3.

**Tabel 4.3** Hasil Fitur Minimum Masing-Masing *Sub-band*

<b>Minimum</b>				
<b>Potongan</b>	<b>Theta</b>	<b>Alpha</b>	<b>Beta</b>	<b>Label</b>
<b>0</b>	-39,5327	-26,5202	-26,4844	0
<b>1</b>	-255,916	-165,356	-137,831	0
<b>2</b>	-87,2386	-92,6702	-80,0322	0
<b>3</b>	-74,3964	-32,7649	-27,2532	0
<b>4</b>	-126,592	-95,6296	-140,659	0
<b>:</b>	<b>:</b>	<b>:</b>	<b>:</b>	<b>:</b>
<b>498</b>	-58,8063	-75,7439	-70,5682	0
<b>499</b>	-68,034	-97,7054	-74,6302	0

Selanjutnya dengan menggunakan data pada Tabel 4.3 akan dibuat grafik dari *sub-band theta*, *alpha* dan *beta* untuk melihat perbedaan dari fitur minimum untuk sinyal kejang dan sinyal non-kejang yang ditampilkan pada Gambar 4.6.



**Gambar 4.6** Plot Nilai Minimum Gelombang *Theta* (a), *Alpha* (b), dan *Beta* (c)

Pada Gambar 4.6 dapat dilihat bahwa plot sinyal kejang (label S) memiliki nilai yang cenderung berada dibawah kelompok sinyal non-kejang untuk ketiga gelombang. Sama halnya dengan

fitur sebelumnya, fitur minimum untuk gelombang *theta* dan *beta* memberikan perbedaan yang signifikan antar sinyal kejang-dan sinyal non-kejang. Namun untuk gelombang *alpha* terdapat cukup banyak nilai dari kelompok sinyal label O yang masuk kedalam *range* nilai amplitudo sinyal kejang.

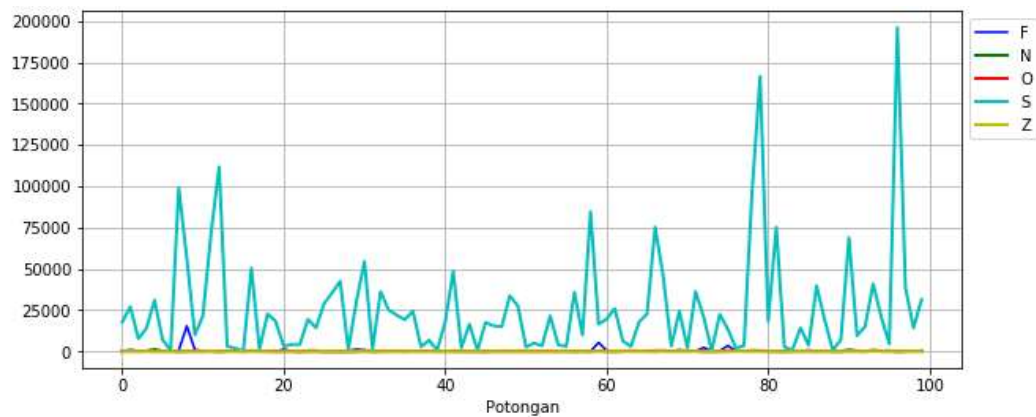
#### 4.3.3 Fitur *Variance*

Fitur *variance* merupakan fitur yang mengukur seberapa besar persebaran nilai sampel dari masing-masing *sub-band* yang dihitung dengan menggunakan Persamaan 2.4. Pada fitur *variance* ini akan dihasilkan 3 variabel yang akan digunakan sebagai variabel prediktor untuk membedakan antara sinyal kejang dan sinyal non-kejang. Adapun hasil perhitungan dari fitur *variance* akan ditunjukkan pada Tabel 4.4.

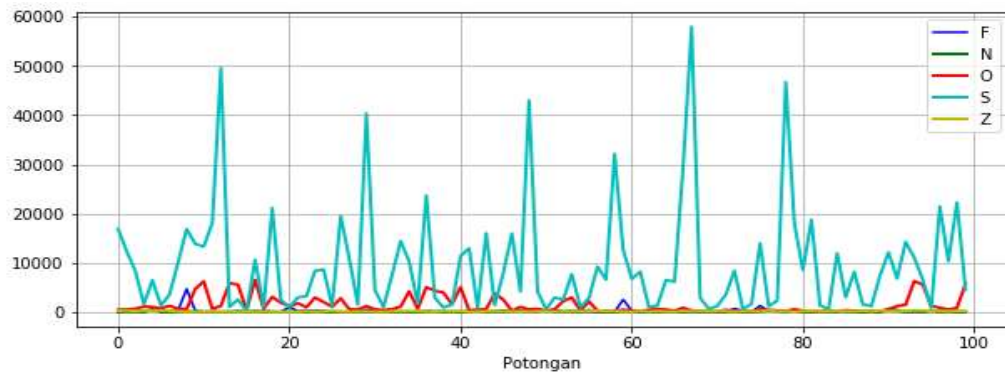
**Tabel 4.4** Hasil Fitur *Variance* Masing-Masing *Sub-band*

<i>Variance</i>				
Potongan	Theta	Alpha	Beta	Label
0	119,0749	44,44404	31,76824	0
1	1984,413	694,5579	327,6349	0
2	672,3841	663,6442	216,4214	0
3	381,1037	79,98876	35,90848	0
4	1512,853	856,5264	569,551	0
⋮	⋮	⋮	⋮	⋮
498	376,9031	528,792	532,2251	0
499	255,2569	476,3428	323,8086	0

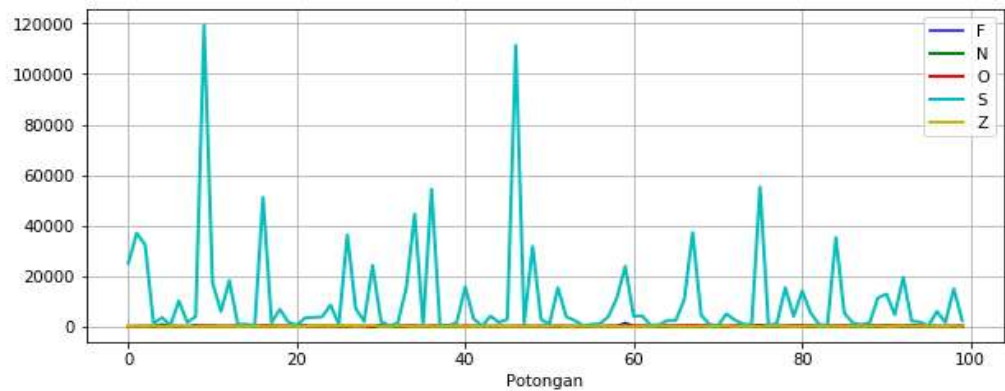
Untuk melihat lebih jelas perbedaan antara sinyal kejang dan sinyal non-kejang dengan menggunakan fitur *variance* akan ditampilkan grafik dari setiap satu *sub-band*.



(a)



(c)



(c)

**Gambar 4.7** Plot Nilai *Variance* Gelombang *Theta* (a), *Alpha* (b), dan *Beta* (c)

Gambar 4.7 merupakan plot nilai *Variance* gelombang *theta*, *alpha*, dan *beta* untuk kelima kelompok sinyal. Dari gambar 4.6 dapat dilihat bahwa nilai *variance* untuk sinyal kejang memiliki nilai yang lebih besar dan juga lebih fluktuatif dari kelompok sinyal

lainnya. Sehingga fitur ini akan dapat digunakan untuk membedakan antara sinyal kejang dan sinyal non-kejang. Berdasarkan ketiga gambar dapat disimpulkan bahwa sinyal dengan hasil fitur ekstraksi *variance* untuk gelombang *alpha* cenderung lebih sulit membedakan sinyal kelompok kejang dengan sinyal non-kejang jika dibandingkan dengan gelombang *theta* dan *beta*.

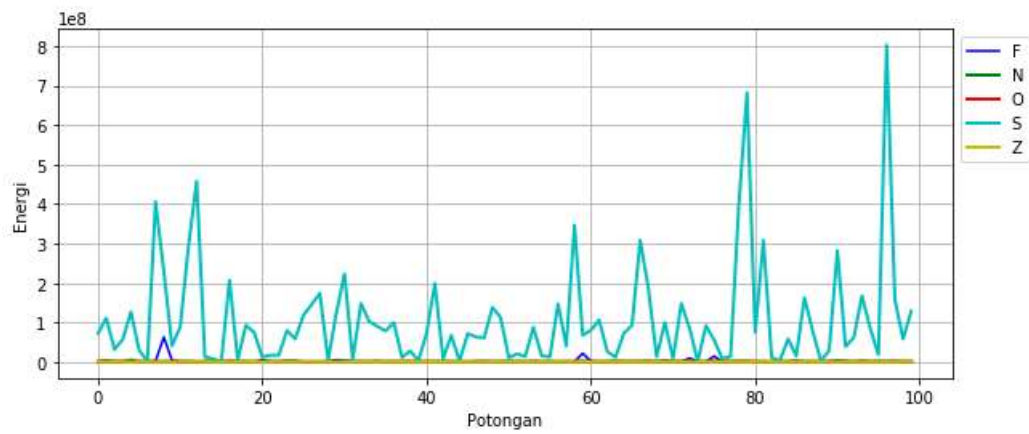
#### 4.3.4 Fitur Energi

Energi dari sinyal EEG dapat mengandung informasi yang berguna mengenai sifat penyakit yang memengaruhi otak. Fitur energi dihitung untuk masing-masing *subband theta*, *alpha*, dan *beta* dengan menggunakan Persamaan 2.3. Dari fitur energi akan dihasilkan matriks dengan ukuran  $3 \times 500$ , dengan nilai 500 menunjukkan jumlah dataset. Adapun hasil perhitungan nilai energi akan ditampilkan pada Tabel 4.5.

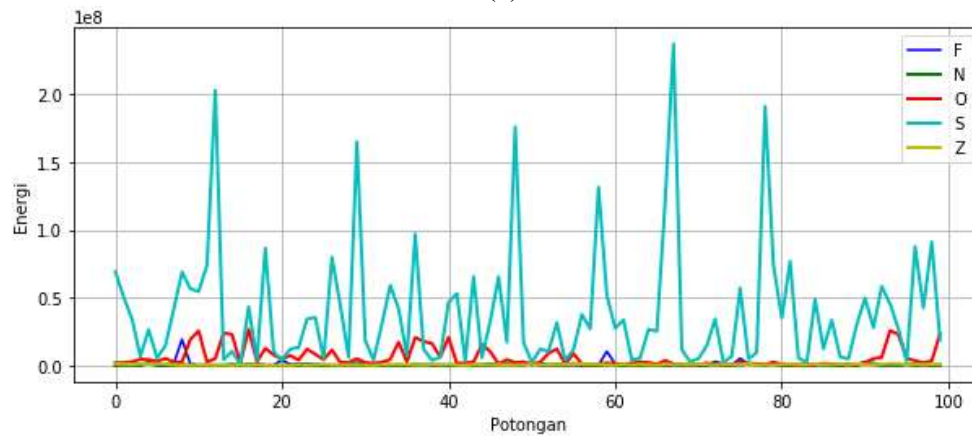
**Tabel 4.5** Hasil Fitur Energi Masing-Masing *Sub-band*

Potongan	Energi				Label
	Theta	Alpha	Beta		
0	$4,88 \times 10^5$	$1,82 \times 10^5$	$1,30 \times 10^5$		0
1	$8,13 \times 10^6$	$2,85 \times 10^6$	$1,34 \times 10^6$		0
2	$2,75 \times 10^6$	$2,72 \times 10^6$	$8,87 \times 10^5$		0
3	$1,56 \times 10^6$	$3,28 \times 10^5$	$1,47 \times 10^5$		0
4	$6,20 \times 10^6$	$3,51 \times 10^6$	$2,33 \times 10^6$		0
⋮	⋮	⋮	⋮		⋮
498	$1,54 \times 10^6$	$2,17 \times 10^6$	$2,18 \times 10^6$		0
499	$1,05 \times 10^6$	$1,95 \times 10^6$	$1,33 \times 10^6$		0

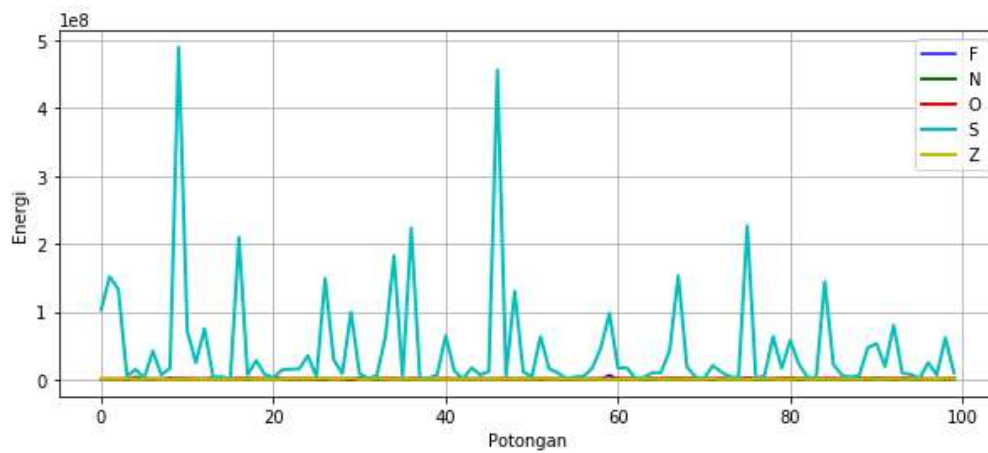
Berikut merupakan *line plot* dari fitur energi untuk setiap *sub-band* yang terdapat pada Tabel 4.5.



(a)



(b)



(b)

**Gambar 4.8** Plot Nilai Energi Gelombang *Theta* (a), *Alpha* (b), dan *Beta* (c)

Gambar 4.8 menunjukkan nilai energi untuk kelompok sinyal kejang cenderung berada jauh diatas kelompok sinyal non-kejang. Hasil ekstraksi fitur energi memberikan kesimpulan yang sama dengan fitur-fitur sebelumnya, dimana fitur untuk gelombang *beta* memiliki banyak nilai yang beririsan antara sinyal kejang dan sinyal non-kejang.

#### 4.3.5 Fitur *Entropy*

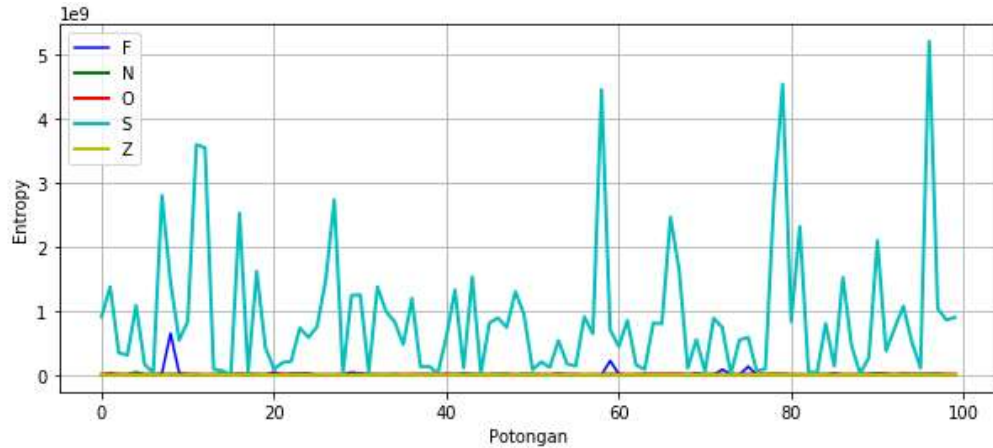
*Entropy* merupakan salah satu fitur yang sering digunakan dalam menganalisis sinyal biomedis. *Entropy* dihitung pada masing-masing *sub-band* dengan menggunakan Persamaan 2.5. Pada fitur *entropy* akan akan dihasilkan jumlah fitur sebanyak 3 fitur seperti yang ditunjukkan pada Tabel 4.6.

**Tabel 4.6** Hasil Fitur *Entropy* Masing-Masing *Sub-band*

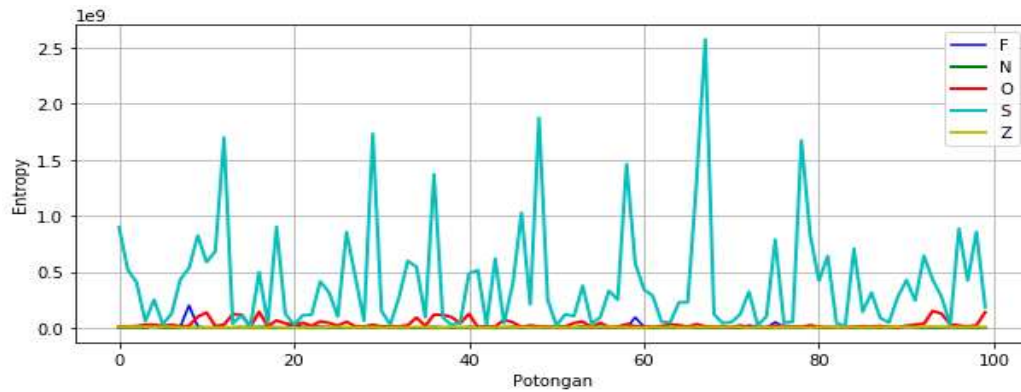
<i>Variance</i>				
Potongan	Theta	Alpha	Beta	Label
0	$1,18 \times 10^6$	$3,60 \times 10^5$	239940,2	0
1	$3,29 \times 10^7$	$1,03 \times 10^7$	4777600	0
2	$8,76 \times 10^6$	$8,76 \times 10^6$	2548496	0
3	$4,62 \times 10^6$	$7,36 \times 10^5$	280414,2	0
4	$2,17 \times 10^7$	$1,15 \times 10^7$	7783464	0
⋮	⋮	⋮	⋮	⋮
498	$4,47 \times 10^6$	$6,65 \times 10^6$	6651643	0
499	$2,87 \times 10^6$	$6,03 \times 10^6$	3758501	0

Pada Gambar 4.9 akan ditampilkan *line plot* hasil ekstraksi fitur *entropy* untuk setiap *sub-band*. Dari Gambar 4.8 dapat dilihat bahwa *line plot* untuk sinyal kejang cenderung berada diatas kelompok sinyal non-kejang. Selain itu sinyal kejang yang ditunjukkan dengan label S nilainya cenderung fluktuatif,

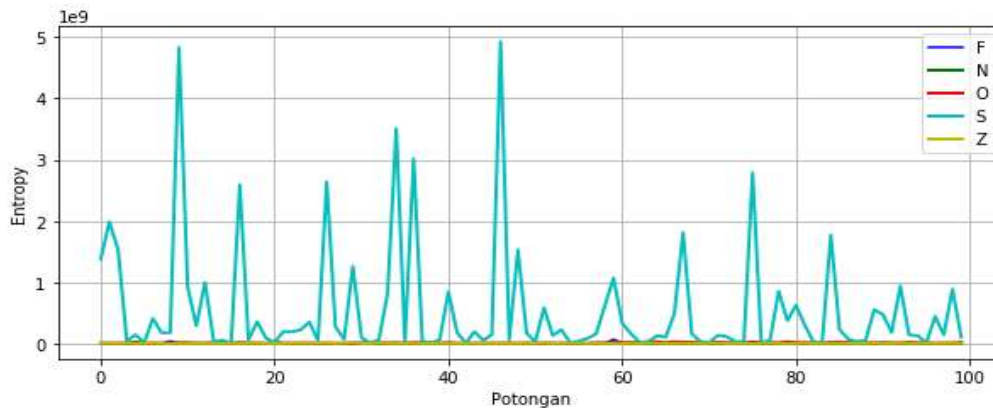
sementara nilai untuk kelompok sinyal non-kejang yaitu label F, N, O, dan Z cenderung konstan.



(a)



(b)



(c)

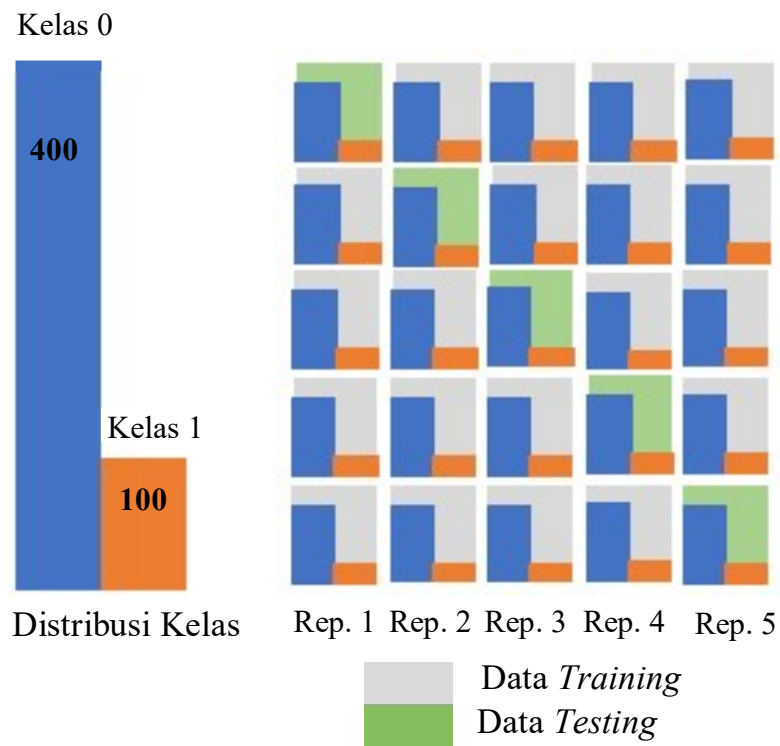
**Gambar 4.9** Plot Nilai *Entropy* Gelombang *Theta* (a), *Alpha* (b), dan *Beta* (c)



Dari Gambar 4.9 dapat dilihat bahwa *range* nilai untuk plot gelombang *alpha* cenderung lebih besar dari pada kelompok lainnya, sehingga nilai fitur *Entropy* gelombang *alpha* untuk sinyal non-kejang yang berasal dari label O cenderung banyak yang beririsan dengan kelompok sinyal kejang.

#### 4.4 Pembagian Data *Traning* dan Data *Testing*

Setelah tahapan ekstraksi fitur selesai, tahapan selanjutnya adalah pembagian data dengan menggunakan metode *stratified 5-fold cross validation*. Pada Gambar 4.10 akan ditampilkan ilustrasi pembagian data *traning* dan *testing* dengan metode *statified 5-fold cross validation*.



**Gambar 4.10** Pembagian Data Dengan Menggunakan *Stratified 5-Fold Cross Validation*

Pada tahap awal data akan dibagi menjadi 5 bagian (*fold*) yang sama dengan proporsi yang sama antara data kejang dan non-

kejang disetiap *fold*-nya. Sehingga dengan menggunakan *dataset* sebanyak 500 data dimana terdapat 100 data sinyal kejang dan 400 data sinyal non-kejang masing-masing *fold* akan beranggotakan 20 data sinyal kejang dan 80 data sinyal non-kejang.

Pemilihan metode ini didasarkan pada penelitian-penelitian sebelumnya yang dilakukan oleh Tsamardinos, Rakhshani, dan Lagani, (2014) yang menyimo bahwa metode *statified k-fold cross-validation* cocok digunakan untuk data yang *imbalance* sehingga akan dihasilkan proporsi data *traning-testing* yang seimbang untuk masing-masing label.

#### 4.5 K-Nearest Neighbor

Setelah melalui tahapan pembagian *traning* dan *testing* data, tahapan berikutnya adalah melakukan klasifikasi menggunakan KNN, langkah awal yang dilakukan adalah mencari jarak terdekat antara data *testing* dengan data *traning* sebanyak *k*. Untuk lebih memperjelas cara kerja KNN ilustrasi perhitungannya ditampilkan pada Tabel 4.7. Data ilustrasi yang digunakan dipilih dari 500 data sinyal dan 15 variabel yang terdapat pada Lampiran 3 dengan 5 data sinyal berasal dari label 0 dan 2 sinyal berasal dari label 1 dan 1 sinyal belum memiliki label.

**Tabel 4.7** Ilustrasi Sampel Data

No	X1	X2	X3	...	X15	Label
1	218492,9	25,24188	-25,9535	...	239940,2	0
2	3949624	133,6558	-129,123	...	4777600	0
3	1283750	62,4215	-61,7414	...	2548496	0
4	649870,3	45,6576	-45,5875	...	280414,2	0
5	3092881	76,09213	-72,9873	...	7783464	0
6	8789756	128,8298	-131,011	...	20504527	1
7	14189509	177,7345	-170,643	...	51101176	1
8	546562,4	41,80466	-43,0403	...	3758501	?

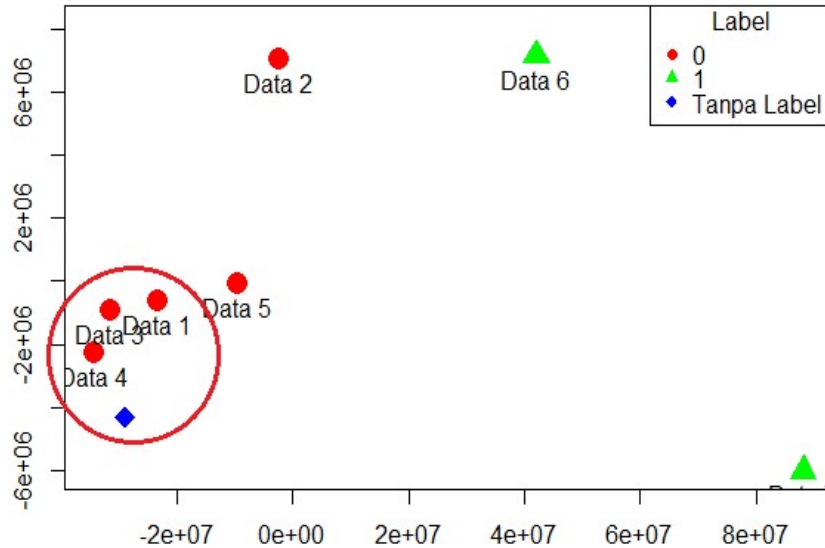
Tabel 4.7 menunjukkan bahwa data ke-8 belum memiliki kelas/label. Untuk menentukan kelas data tersebut, dapat dilakukan dengan menghitung jarak terdekat antara data ke-8 dengan masing-masing data yang telah memiliki label seperti pada Tabel 4.8. Penghitungan jarak data dihitung dengan menggunakan metode *euclidean* seperti pada Persamaan 2.6.

**Tabel 4.8** Jarak *Euclidean* Data ke-8 dengan Data Lainnya

No	X1	X2	...	X15	Jarak	Label
1	218492,9	25,24188	...	239940,2	6962159	0
2	3949624	133,6558	...	4777600	30568998	0
3	1283750	62,4215	...	2548496	6658076	0
4	649870,3	45,6576	...	280414,2	6639576	0
5	3092881	76,09213	...	7783464	20215520	0
6	8789756	128,8298	...	20504527	71933357	1
7	14189509	177,7345	...	51101176	1.17E+08	1

Setelah didapatkan jarak data ke-8 dengan masing-masing data yang telah diketahui labelnya, selanjutnya adalah mengurutkan nilai jarak dari nilai yang terkecil ke yang terbesar dan memilih nilai  $k$ - tetangga terdekat. Dengan nilai  $K = 3$ , maka dengan melihat 3 data sampel dengan jarak terdekat terhadap data ke-8, dapat ditentukan bahwa data ke-8 masuk kedalam label 0. Penentuan ini berdasar kepada mayoritas label dari 3 data dengan jarak terdekat. Untuk memperjelas penentuan kelas, Gambar 4.10 akan memberikan ilustrasi penentuan kelas berdasarkan jarak terdekat.

Jika dicocokkan hasil dari Tabel 4.8 dan ilustrasi pada Gambar 4.11 dapat dilihat saat  $K = 3$  maka akan dicari 3 titik dengan nilai jarak *euclidean* terdekat yaitu data sinyal ke- 4, data sinyal ke-3, dan data sinyal ke- 1 dimana ketiga data sinyal tersebut memiliki label 0. Sehingga data sinyanya yang belum memiliki label akan diklasifikasikan sebagai kelompok data sinyal label 0.



**Gambar 4.11** Ilustrasi Penentuan Kelas Berdasarkan Mayoritas Kelas Jarak Terdekat

Penentuan pada jarak terdekat dengan metode *K-Nearest Neighbor* ini berdasarkan pada tinjauan pustaka pada buku karangan Han, Kamber, & Pei (2012). Pada tugas akhir ini akan digunakan nilai  $K = 1$  sampai dengan  $K = 19$  untuk melihat pengaruh nilai  $k$  yang memberikan performansi pengklasifikasian terbaik. Jarak yang digunakan dalam penelitian ini adalah dengan menggunakan jarak *euclidean* yang ditunjukkan pada Lampiran 4. Berikut merupakan hasil performansi klasifikasinya.

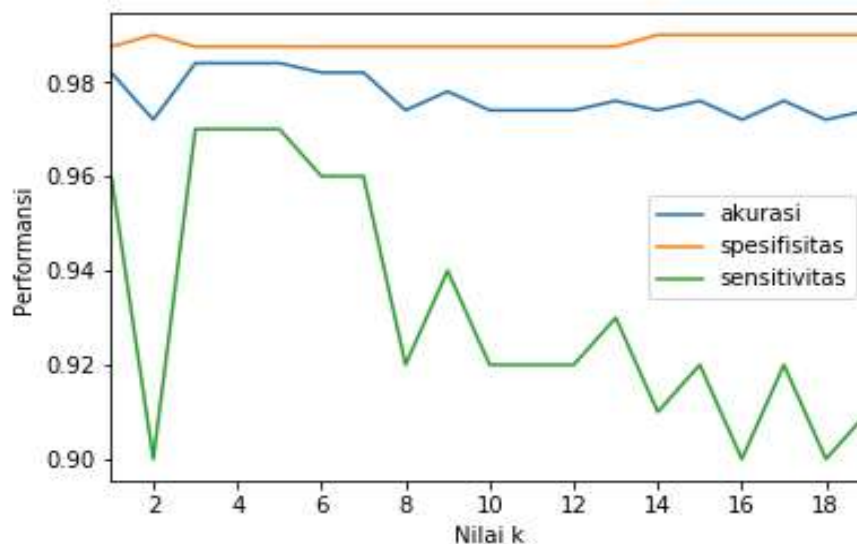
**Tabel 4.9** Perbandingan Performansi Klasifikasi Dengan Nilai K yang Berbeda

K-	Akurasi	Sensitivitas	Spesifisitas
1	0,982	0,96	0,9875
2	0,97	0,89	0,99
3	0,984	0,97	0,9875
4	0,984	0,97	0,9875
5	0,984	0,97	0,9875
6	0,982	0,96	0,9875
7	0,982	0,96	0,9875
8	0,976	0,93	0,9875

**Tabel 4.10** Perbandingan Performansi Klasifikasi Dengan Nilai K yang Berbeda (Lanjutan)

K-	Akurasi	Sensitivitas	Spesifisitas
9	0,978	0,94	0,9875
10	0,974	0,92	0,9875
11	0,974	0,92	0,9875
12	0,974	0,92	0,9875
13	0,974	0,92	0,9875
14	0,976	0,92	0,99
15	0,976	0,92	0,99
16	0,972	0,9	0,99
17	0,976	0,92	0,99
18	0,972	0,9	0,99
19	0,976	0,92	0,99

Tabel 4.9 menunjukkan bahwa nilai akurasi paling optimal adalah saat  $K = 3, 4$ , dan  $5$  dengan nilai akurasi sebesar  $0,984$ . Nilai sensitivitas paling optimal adalah saat nilai  $K = 3, 4$  dan  $5$ . Sedangkan untuk nilai spesifisitas optimum pada saat nilai  $k = 2, 14, 15, 16, 17, 18$  dan  $19$ . Untuk lebih jelasnya grafik perbandingannya akan ditunjukkan pada Gambar 4.12.



**Gambar 4.12** Grafik Perbandingan Performansi Nilai K

Gambar 4.12 menunjukkan bahwa nilai sensitivitas dan akurasi cenderung turun setelah nilai  $k = 5$ , namun nilai sensitivitas cenderung konstan disetiap nilai  $k$ . Secara umum dapat disimpulkan bahwa nilai  $k$  terbaik untuk data set tugas akhir ini adalah saat nilai  $k=3, 4$ , dan  $5$ . Selain itu nilai  $k$  yang semakin besar mengakibatkan nilai sensitivitas yang semakin turun.

Pada tugas akhir ini, nilai  $k$  tetangga terdekat yang digunakan adalah saat  $k=3$ . Pemilihan nilai  $k = 3$  berdasar kepada beban komputasi, dimana beban komputasi untuk  $k = 4$  atau  $k = 5$  cenderung lebih besar jika dibandingkan dengan nilai  $k = 3$ . Dengan menggunakan nilai  $k = 3$  didapatkan nilai akurasi sebesar 98,4%. Artinya dari 500 data, data sinyal yang mengalami misklasifikasi adalah sebesar 8, untuk melihat rincian data sinyal yang salah diklasifikasi dapat dilihat pada Tabel 4.10.

**Tabel 4.11** *Confusion Matrix* Klasifikasi Terbaik dengan Metode 3-NN

		Prediksi		
		Kejang	Non-Kejang	Total
Aktual	Kejang	97	3	100
	Non-Kejang	5	395	400
	Total	102	398	500

Merujuk pada Tabel 4.10 diketahui total misklasifikasi adalah sebanyak 8 data sinyal. Dengan rincian data sinyal non-kejang yang salah diklasifikasikan menjadi data sinyal kejang ada sebanyak 5 dari total 400 data sinyal atau kesalahan klasifikasi sebesar 1,25% untuk kelas sinyal non-kejang, serta terdapat 3 dari 100 data sinyal kejang yang salah diklasifikasikan sebagai data sinyal non-kejang atau ada kesalahan klasifikasi sebesar 3% untuk kelas sinyal kejang.

Pada Tabel 4.10 didapatkan nilai sensitivitas dan spesifisitas yang mendekati 100%, dan nilai paling optimum diberikan saat  $k = 3, 4, 5, 6$ , dan  $7$ . Saat  $k = 3$  didapatkan nilai

sensitivitas adalah sebesar 97%, artinya hasil yang diberikan oleh sistem mampu mengklasifikasikan data sinyal kejang dengan ketepatan 97%. Sedangkan untuk nilai spesifisitas adalah sebesar 98,75% yang artinya sistem mampu mengklasifikasikan sinyal non-kejang dengan ketepatan sebesar 98,75%.

Berdasarkan nilai sensitivitas dan spesifisitas dapat dihitung akurasi klasifikasi untuk data *imbalance*. Akurasi klasifikasi yang digunakan adalah *G-mean* dan *Area Under ROC Curve* (AUC). Perhitungan akurasi menggunakan kriteria *G-mean* akan sangat sensitif apabila jumlah data *testing* pada salah satu kategori sangat kecil, maka untuk mengatasi permasalahan tersebut dilakukan pula perhitungan akurasi menggunakan kriteria AUC. Perlu diketahui bahwa pada penelitian ini terdapat 100 data sinyal kejang sementara untuk data sinyal non-kejang terdapat sejumlah 400 data sinyal, sehingga pada data *testing* akan terdapat jumlah data kejang yang sangat sedikit jika dibandingkan dengan data non-kejang.

Perhitungan akurasi menggunakan *G-mean* dapat dihitung dengan menggunakan Persamaan 2.11.

$$GM = \sqrt{0,97 \times 0,9875} = 97,25\%$$

Sedangkan untuk perhitungan akurasi menggunakan kriteria AUC dapat dihitung dengan menggunakan Persamaan 2.12.

$$AUC = \frac{1}{2} (0,97 + 0,9875) = 97,87\%$$

Bedasarkan perhitungan akurasi dengan menggunakan *G-mean* dan AUC didapatkan bahwa hasil klasifikasi sinyal EEG menggunakan KNN memiliki hasil yang sangat baik yaitu dengan nilai sebesar 97,25% dan 97,87%.

Hasil pembahasan sebelumnya diketahui bahwa metode 3-NN memberikan hasil klasifikasi yang terbaik dengan eror terkecil yaitu sebesar 1,6% dengan ketepatan klasifikasi sebesar 98,4%. Nilai pada hasil tersebut merupakan nilai yang didapatkan dari rata-rata keseluruhan semua *fold*. Perlu diketahui bahwa setiap *fold* memiliki jumlah data sebanyak 100 dengan rincian 20 data sinyal kejang dan 100 data sinyal non-kejang. Hasil pengukuran performansi klasifikasi tiap *fold* ditunjukkan pada Tabel 4.11.

**Tabel 4.12** Perbandingan Performansi Klasifikasi dengan  $k=3$  Pada Masing-Masing *Fold*

<b>Fold</b>	<b>Akurasi</b>	<b>Sensitivitas</b>	<b>Spesitifitas</b>
1	0,99	1	0,9875
2	0,99	1	0,9875
3	0,98	0,95	0,9875
4	0,99	1	0,9875
5	0,97	0,9	0,9875
<b>Rata-rata</b>	0,984	0,97	0,9875
<b>Standar Deviasi</b>	0,008944	0,044721	0

Dari Tabel 4.11 diketahui bahwa dari 5 *fold* yang digunakan *fold* yang memberikan nilai akurasi dan sensitivitas terbaik adalah *fold* ke 1, 2, dan 4 yaitu dengan akurasi dan sensitivitas masing-masing sebesar 99% dan 100%, yang artinya pada *fold* tersebut sistem mampu mengklasifikasikan sinyal kejang dan non-kejang sebesar 99% atau dengan kata lain hanya satu observasi yang salah diklasifikasikan. Sementara nilai sensitivitas sebesar 100% menunjukkan bahwa sistem mampu mengklasifikasikan semua data sinyal kejang dengan benar.

Berdasarkan performansi dari nilai spesitifitas semua *fold* menghasilkan nilai yang sama yaitu sebesar 98,75%. Artinya bahwa sistem mampu mengklasifikasikan data sinyal non-kejang dengan ketepatan 98,75%, dimana dari 80 sinyal kejang yang digunakan sebagai data *testing* hanya terdapat 1 sinyal yang salah diklasifikasikan sebagai sinyal non-kejang.



## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Hasil filterisasi sinyal terdiri atas 3 gelombang yaitu *theta*, *alpha*, dan *beta*, masing-masing gelombang memberikan perbedaan yang signifikan antara sinyal kejang dan sinyal non-kejang. Kemudian dari hasil ekstraksi fitur sinyal dengan gelombang *theta* dan *beta* memberikan perbedaan yang signifikan antara sinyal kejang dan sinyal non-kejang. Sementara sinyal dengan gelombang *alpha* kurang signifikan dalam membedakan sinyal kejang dan sinyal non-kejang.
2. Pengklasifikasian signal EEG dengan menggunakan metode KNN didapatkan nilai K terbaik adalah saat  $K = 3$  dengan akurasi, sensitivitas, dan spesitifitas masing-masing sebesar 98,4%, 97,0%, dan 98,75%. Selanjutnya dengan menggunakan nilai tersebut didapatkan nilai *G-mean* dan AUC sebesar 97,25% dan 97,87%.

#### **5.2 Saran**

Berdasarkan kesimpulan yang diperoleh, dapat dirumuskan saran sebagai pertimbangan penelitian selanjutnya adalah sebagai berikut.

1. Pada Metode filterisasi dengan metode *band-pass butterworth* gelombang atau *sub-band* yang difilter cukup gelombang *theta* dan *beta*.
2. Pada metode KNN, sebaiknya menggunakan metode optimasi parameter untuk meningkatkan performansi hasil klasifikasi.

*(Halaman ini sengaja dikosongkan)*

## DAFTAR PUSTAKA

- Alotaiby, F. E., El-Samie, F. E., Alshebeili, S. A., Aljibreen, K. H., Alkhanen, E. (2015). Seizure Detection with Common Spatial Pattern and Support Vector Machines. *International Conference on Information and Communication Technology Research (ICTRC2015)*.
- Andarini, I. (2007). Hubungan Kepatuhan Pengobatan dengan Remisi Epilepsi pada Anak. *Laporan Penelitian Akhir*. IP Saraf FK UGM. Yogyakarta.
- Beatty, J. (2001). *The Human Brain: Essentials of Behavioral Neuroscience*, Thousand Oak, CA : Sage Publicaion.
- Bekkar, M., Djemaa, H. K., & Alitouch, T. A. (2013). Evaluation Measure for Models Assesment over Imbalanced Data Sets. *Journal of Information Engineering and Aplications*, 3, 27-38.
- Bhavsar, H and Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering (IJSCE)*. vol. 2, no. 4, Sep.
- Brodie M.J., Schachter S.C, Kwan P. (2012). *Epidemiology and Prognosis*. Fast Fact: Epilepsy Revised 5th edition. Oxford: Health Press Limited. pp: 9-11.
- Campellone, J.V. *EEG BRAIN WAVE TEST* Diambil pada 19 Februari 2018 dari <http://www.nlm.nih.gov/medlineplus/ency/article/003931.htm>.
- Duda, R. and Hart, P. (1973). *Pattern Classification and Scene Analysis*. John Wiley and Sons.
- Engel J, Pedley T. (2008). *Epilepsy, a comprehensive text book*. 2nd ed. Philadelphia: Lippincot, William and Willkins. p 219-89.

- Guo, L., Rivero, D., Dorado, J., Munteanu, C.R., Pazos, A. (2011). Automatic feature extraction using genetic programming: an application to epileptic EEG classification. *Expert Systems with Applications*. 38 (8) 10425–10436.
- Hamad, A. Houssein, E. H., Hassanien, A. E., Fahmy, A. A., (2016). Feature Extraction of Epilepsy EEG using Discrete Wavelet Transform. *Computer Engineering Conference (ICENCO)*.
- Han, J., Kamber, M., dan Pei, J. (2012). *DATA MINING. Concepts and Techniques*. Third Edition. Morgan Kaufmann Publishers.
- Harsono. (1999). *Buku Ajar Neurologi Klinis*. Jakarta, Dian Rakyat.
- Hotho, A., Nurnberger, A., & Paass, G. (2005). *A Brief Survey of Text Mining*. Kassel: University of Kassel.
- Karlık, B., dan Hayta, S. B. (2014). Comparison Machine Learning Algorithms for Recognition of Epileptic Seizures in EEG. *Proceedings IWBBIO*. Granada 7-9 Apri.
- Korba, H. B. M.C. Amara. Application of Empirical Mode Decomposition and Artificial Neural Network For The Classification of Normal and Epileptic EEG Signal. *Biocybernetics and Biomedical Engineering*. Vol.36, Issue 1, pp.285-291.
- Krishnan, M., Neophytou C.P., Prescott G. (1994). Wavelet Transform Speech Recognition Using Vector Quantization, Dynamic Time Warping and Artificial Neural Networks. Center of excellence in computer aided systems engineering and Telecommunication & Information Sciences Laboratory 2291 Irving Hill Drive. Lawrence, KS 66045.
- Kumar, A. & Kolekar, M. H. (2014). Machine Learning Approach for Epileptic Seizure Detection Using Wavelet Analysis of EEG Signals. *International Conference on Medical*

*Imaging, m-Health and Emerging Communication system (MedCom).* pp 412-416.

- Kusumastuti, K., Gunadharma, S., dan Kustiowati E. (2011). *Pedoman Tatalaksana Epilepsi Kelompok Studi Epilepsi Perhimpunan Dokter Spesialis Saraf Indonesia (PERDOSSI)*, Airlangga University Press, Surabaya.
- Napitupulu, L. Y., Suciati, N., dan Navastara, D. A. (2017). Implementasi Deteksi Serangan Epilepsi dari Data Rekaman EEG Menggunakan Weighted Permutation Entropy dan Support Vector Machine. *JURNAL TEKNIK ITS* Vol. 6, No. 2.
- Nguyen, A. Thanh. (2015). EEG Signal Clasification for BCI applications by Wavelets and Interval Type-2 Fuzzy Logic Systems. *Expert Systems with Applications*. p.4370-4380.
- Podder, P., Hasan, M., Islam M., Sayeed, M. (2014). Design and Implementation of Butterworth, Chebyshev-I and Elliptic Filter for Speech Signal Analysis. *International Journal of Computer Applications (0975 – 8887)*. Vol. 98– No.7, July 2014.
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta: Andi Publisher.
- Shoeb, A., dan Guttag, J. (2010). Application of Machine Learning to Epileptic Seizure Detection. *International Conference on Machine Learning*.
- Smith, Steven W. (1997). *The Scientist and Engineer's Guide to Digital Signal Processing*. California, USA: California Technical Publishing.
- Soliman, Samir S. dan Mandyam D. Srinath. (1990). *Continuous and Discrete Signals and Systems*. New Jersey, USA: Prentice-Hall Inc.
- Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D., dan Levy, S. (2005). A Comprehensive Evaluation Of

- Multicategory Classification Methods For Microarray Gene Expression Cancer Diagnosis, *Bioinformatics*, vol. 21, pp. 631-643.
- Subasi, A. (2013). Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Computers in Biology and Medicine*, vol. 43, pp 576-586.
- Subha, D. Puthankattil, Paul K. Joseph, Rajendra Acharya U. and Lim, Choo Min. EEG Signal Analysis: A Survey. *Journal of Medical System*. Vol.34 (2),pp 195-212,2010.
- Sun, Y., Kamel, M. S., & Wang, Y. (2006). Boosting for Learning Multiple Classes with Im-balanced Class Distribution. *Sixth International Conference on Data Mining (ICDM'06)*, 421-431.
- Tamouk, J., and Allahakbari, F. (2012). A comparison among accuracy of KNN, PNN, KNCN, DANN and NFL. *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 3 No. 1, pp. 319-322.
- Teplan, M. 2012. Fundamentals of EEG Measurement. *Measurement Science Review*. vol. 2, no. 2, pp. 1–11.
- Tsamardinos, I., Rakhshani1, A., & Lagan, V. (2015). Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization. *International Journal on Artificial Intelligence Tools*. Vol. 24.
- World Health Organization. (2018). Diambil pada 18 Februari 2018 dari <http://www.who.int/mediacentre/factsheets/fs999/en/>.
- Yang, S., Huang, J., Zhou, D., Zha, H., and Giles, C. L. (2007) IKNN: Informative KNearest Neighbor Pattern Classification. *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Warsaw, Poland, Springer-Verlag. pp. 248-264.

## LAMPIRAN

### Lampiran 1. Data Sinyal EEG untuk Deskripsi Dataset dan Filterisasi Sinyal

Set	Sampel Ke-				
		1	2	...	4097
<b>F</b>	1	34	33	...	7
	2	60	47	...	42
	⋮				
	100	-45	-53	...	-130
<b>N</b>	1	-42	-39	...	-13
	2	50	48	...	1
	⋮				
	100	-62	-64	...	5
<b>O</b>	1	-24	-22	...	-31
	2	-55	-48	...	-51
	⋮				
	100	-18	-13	...	-39
<b>S</b>	1	100	124	...	119
	2	340	353	...	9
	⋮				
	100	23	144	...	910
<b>Z</b>	1	12	22	...	-212
	2	-56	-50	...	217
	⋮				
	100	-36	-71	...	82

**Lampiran 2.** Data Hasil Filterisasi Sinyal EEG

<b>Amplitudo Ke-</b>	<b>Theta F1</b>	<b>Theta F2</b>	<b>...</b>	<b>Beta Z99</b>	<b>Beta Z100</b>
<b>1</b>	0.000004	0.000007		0.022944	-0.01475
<b>2</b>	0.000045	0.000079		0.18912	-0.13618
<b>3</b>	0.000274	0.000469		0.662446	-0.57104
<b>4</b>	0.001104	0.001857		1.151932	-1.39519
<b>5</b>	0.003367	0.005572		0.433205	-1.99739
<b>6</b>	0.008359	0.01362		-2.71939	-0.91452
<b>7</b>	0.017696	0.028401		-7.12427	3.086478
<b>8</b>	0.032958	0.052136		-8.18041	9.043839
<b>9</b>	0.055198	0.086092		-1.16196	12.86393
<b>10</b>	0.084382	0.129792		12.90653	9.249584
<b>⋮</b>	<b>⋮</b>	<b>⋮</b>		<b>⋮</b>	<b>⋮</b>
<b>4095</b>	-13.2179	81.77098		-19.6374	21.52795
<b>4096</b>	-13.2463	77.19127		-39.4103	7.462987
<b>4096</b>	-12.8012	68.46351		-26.611	-5.47978



### Lampiran 3. Data Hasil Ekstraksi Fitur

#### Fitur Maksimum

	<b>Theta</b>	<b>Alpha</b>	<b>Beta</b>	<b>Kelas</b>
<b>1</b>	39.39205	23.50233	30.86392	0
<b>2</b>	257.4043	141.8635	127.7205	0
<b>3</b>	87.46132	83.96361	83.10808	0
<b>4</b>	70.59756	29.16685	25.75352	0
<b>5</b>	128.5275	93.37347	151.2898	0
<b>6</b>	50.41843	25.01578	16.9073	0
<b>7</b>	88.56467	53.44998	43.45684	0
<b>8</b>	212.0497	128.9623	104.082	0
<b>9</b>	624.3537	397.373	439.3024	0
<b>10</b>	253.3513	156.6187	137.3516	0
<b>11</b>	91.77961	46.70087	43.94788	0
<b>12</b>	74.00316	40.0817	42.26074	0
<b>13</b>	44.92723	25.50634	22.59018	0
<b>14</b>	49.55179	27.98702	19.32341	0
<b>15</b>	70.5637	28.65452	17.5007	0
<b>16</b>	117.0781	60.52178	56.3557	0
<b>17</b>	69.4095	30.88924	19.3002	0
<b>18</b>	122.1727	128.3357	77.2708	0
<b>19</b>	72.36564	56.12971	101.2764	0
<b>20</b>	65.50271	43.44992	41.54023	0
<b>⋮</b>				
<b>499</b>	65.88392	71.36289	84.38345	0
<b>500</b>	60.08065	94.34005	83.48176	0

## Fitur Minimum

	<b>Theta</b>	<b>Alpha</b>	<b>Beta</b>	<b>Kelas</b>
<b>1</b>	-45.5681	-22.3921	-28.2612	0
<b>2</b>	-210.451	-144.162	-142.09	0
<b>3</b>	-95.9211	-80.7101	-82.0605	0
<b>4</b>	-79.5386	-31.5224	-28.183	0
<b>5</b>	-117.278	-96.957	-156.344	0
<b>6</b>	-51.7431	-24.6522	-16.5154	0
<b>7</b>	-94.4307	-52.3259	-48.879	0
<b>8</b>	-195.34	-133.364	-124.666	0
<b>9</b>	-570.557	-404.413	-439.904	0
<b>10</b>	-225.481	-155.223	-150.009	0
<b>11</b>	-93.7005	-50.6726	-42.0593	0
<b>12</b>	-84.4322	-40.55	-37.6429	0
<b>13</b>	-46.8613	-24.6642	-21.6842	0
<b>14</b>	-51.4187	-28.2209	-18.3557	0
<b>15</b>	-66.7411	-29.8379	-15.4971	0
<b>16</b>	-100.409	-62.3738	-67.8523	0
<b>17</b>	-73.6817	-33.1577	-16.2613	0
<b>18</b>	-115.664	-122.569	-67.4767	0
<b>19</b>	-75.2565	-53.1071	-95.7209	0
<b>20</b>	-61.8333	-41.9457	-35.1159	0
<b>⋮</b>				
<b>499</b>	-63.8012	-66.261	-87.2946	0
<b>500</b>	-67.9396	-92.7891	-78.0856	0

Fitur *Variance*

	Theta	Alpha	Beta	Kelas
<b>1</b>	118.5156	35.94751	39.03194	0
<b>2</b>	1976.226	614.1845	395.8575	0
<b>3</b>	664.1757	600.5402	278.9579	0
<b>4</b>	377.3057	69.9769	43.56963	0
<b>5</b>	1498.732	766.0455	648.3962	0
<b>6</b>	232.0274	38.1605	18.2504	0
<b>7</b>	339.3568	171.821	140.6859	0
<b>8</b>	1445.629	533.9211	227.8462	0
<b>9</b>	30969.43	9321.999	3084.291	0
<b>10</b>	2178.486	557.4925	510.6095	0
<b>11</b>	713.4643	143.973	67.75483	0
<b>12</b>	617.4989	113.5577	70.07456	0
<b>13</b>	184.9592	48.69484	30.2277	0
<b>14</b>	305.4929	58.71499	27.0437	0
<b>15</b>	354.1418	62.75361	22.95263	0
<b>16</b>	638.6672	269.3311	202.5378	0
<b>17</b>	635.2021	112.7743	24.33126	0
<b>18</b>	1343.546	927.4047	226.0862	0
<b>19</b>	469.9471	95.67962	58.92068	0
<b>20</b>	184.2524	79.58707	45.76049	0
<b>499</b>	375.2905	390.9386	633.7528	0
<b>500</b>	253.6694	359.9021	427.8885	0

## Fitur Energi

	<b>Theta</b>	<b>Alpha</b>	<b>Beta</b>	<b>Kelas</b>
<b>1</b>	485558.8	147277.3	159913.8	0
<b>2</b>	8096613	2516319	1621833	0
<b>3</b>	2721136	2460413	1142891	0
<b>4</b>	1545834	286695.4	178504.8	0
<b>5</b>	6140307	3138494	2656480	0
<b>6</b>	950620.2	156343.6	74771.89	0
<b>7</b>	1390346	703950.7	576390.4	0
<b>8</b>	5922743	2187475	933485.8	0
<b>9</b>	1.27E+08	38192240	12636340	0
<b>10</b>	8925259	2284047	2091967	0
<b>11</b>	2923091	589857.2	277591.6	0
<b>12</b>	2529894	465245.9	287095.5	0
<b>13</b>	757778.5	199502.8	123842.9	0
<b>14</b>	1251604	240555.3	110798	0
<b>15</b>	1450926	257101.6	94036.92	0
<b>16</b>	2616623	1103449	829797.4	0
<b>17</b>	2602424	462036.2	99685.17	0
<b>18</b>	5504515	3799580	926275.4	0
<b>19</b>	1925374	391999.4	241398	0
<b>20</b>	754882.6	326068.2	187480.8	0
<b>499</b>	1537566	1601680	2596486	0
<b>500</b>	1039284	1474519	1753060	0

Fitur *Entropy*

	<b>Theta</b>	<b>Alpha</b>	<b>Beta</b>	<b>Kelas</b>
<b>1</b>	1178753	359966.4	239940.2	0
<b>2</b>	32920840	10334930	4777600	0
<b>3</b>	8761087	8756340	2548496	0
<b>4</b>	4624013	736341.1	280414.2	0
<b>5</b>	21701550	11532560	7783464	0
<b>6</b>	2517284	332688.5	95691.26	0
<b>7</b>	4126830	2040296	1212131	0
<b>8</b>	22859450	8706249	2227223	0
<b>9</b>	6.51E+08	2E+08	39394132	0
<b>10</b>	35918740	9560297	6361467	0
<b>11</b>	9364336	1712755	497873.5	0
<b>12</b>	7774126	1282941	512561	0
<b>13</b>	1961222	467288.3	182013.5	0
<b>14</b>	3467782	591301.2	156213.6	0
<b>15</b>	4137754	586103.6	129510.8	0
<b>16</b>	8417666	3412160	1855986	0
<b>17</b>	7969612	1326605	132128.4	0
<b>18</b>	18976880	13471490	2053361	0
<b>19</b>	5820996	1145818	476290.6	0
<b>20</b>	2002268	897387.5	293263	0
<b>499</b>	4468435	6653984	6651643	0
<b>500</b>	2869879	6033137	3758501	0

**Lampiran 4. Perhitungan Jarak Antar Data Sinyal Untuk  
Tahapan Klasifikasi**

<b>Jarak</b>	<b>(X1i – X1j)<sup>2</sup></b>	<b>(X2i – X2j)<sup>2</sup></b>	<b>...</b>	<b>(X15i – X15j)<sup>2</sup></b>	<b>Jarak Euclidean</b>
<b>D(1,1)</b>	1.39E+13	1.18E+04	...	2.06E+13	3.38E+07
<b>D(1,2)</b>	1.13E+12	1.38E+03	...	5.33E+12	1.17E+07
<b>D(1,3)</b>	1.86E+11	4.17E+02	...	1.64E+09	3.49E+06
<b>D(1,4)</b>	8.26E+12	2.59E+03	...	5.69E+13	2.48E+07
<b>D(1,5)</b>	2.36E+11	1.23E+02	...	2.08E+10	1.43E+06
<b>D(1,6)</b>	2.61E+11	1.42E+03	...	9.45E+11	3.59E+06
<b>D(1,7)</b>	5.43E+12	7.00E+03	...	3.95E+12	2.35E+07
<b>D(1,8)</b>	3.99E+15	1.13E+05	...	1.53E+15	6.84E+08
<b>D(1,9)</b>	1.90E+13	1.11E+04	...	3.75E+13	3.67E+07
<b>D(1,10)</b>	4.41E+11	2.81E+02	...	6.65E+10	8.33E+06
<b>D(1,11)</b>	4.98E+11	8.96E+02	...	7.43E+10	6.70E+06
<b>D(1,12)</b>	1.35E+10	7.06E+01	...	3.36E+09	8.01E+05
<b>D(1,13)</b>	1.84E+11	1.17E+02	...	7.01E+09	2.34E+06
<b>D(1,14)</b>	9.04E+11	5.50E+02	...	1.22E+10	3.12E+06
<b>D(1,15)</b>	1.08E+12	2.34E+03	...	2.61E+12	8.11E+06
<b>D(1,16)</b>	7.80E+11	1.01E+03	...	1.16E+10	6.92E+06
<b>D(1,17)</b>	1.79E+12	1.88E+03	...	3.29E+12	2.23E+07
<b>D(1,18)</b>	2.46E+11	1.19E+02	...	5.59E+10	4.74E+06
⋮					
<b>D(498,500)</b>	1.13E+12	1.38E+03	...	5.33E+12	1.17E+07
<b>D(499,500)</b>	1.86E+11	4.17E+02	...	1.64E+09	3.49E+06

## Lampiran 5. Program Python untuk Filterisasi Sinyal

```

Import pandas as pd
Import numpy as np
from scipy.signal import butter, lfilter

def butter_bandpass(lowcut, highcut, fs, order=5):
    nyq = 0.5 * fs
    low = lowcut / nyq
    high = highcut / nyq
    b, a = butter(order, [low, high], btype='band')
    return b, a

def butter_bandpass_filter(data, lowcut, highcut, fs, order=5):
    b, a = butter_bandpass(lowcut, highcut, fs, order=order)
    y = lfilter(b, a, data)
    return y

if __name__ == "__main__":
    import numpy as np
    import matplotlib.pyplot as plt
    from scipy.signal import freqz

    # Sample rate and desired cutoff frequencies (in Hz).
    x = np.array([[4, 8, 12]])
    y = np.array([[8, 12, 30]])
    x = np.asmatrix(x)
    y = np.asmatrix(y)
    lowc=dict()
    highc=dict()
    for i in range(3):
        lowc[i]=x[0,i]

```

```

    highc[i]=y[0,i]
fs = 4097.0/23.6

T = 23.6 # seconds
n = int(T * fs) # total number of samples
t = np.linspace(0, T, n, endpoint=False)
for i in range(3):
    b, a = butter_bandpass(lowc[i], highc[i], fs, order=5)
    w, h = freqz(b, a, worN=2000)
    plt.plot((fs * 0.5 / np.pi) * w, abs(h), label="Filter = %d" %
i)
plt.xlabel('Frequency (Hz)')
plt.ylabel('Gain')
plt.grid(True)
plt.legend(loc='best')
# "Noisy" data. We want to recover the 1.2 Hz signal from
this.
fil=dict()
for i in range(0,500):
    a = df1.iloc[:,i]
    fil[i]=[]
    for j in range(3):
        filtered = butter_bandpass_filter(a, lowc[j], highc[j], fs,
order=6)
        fil[i].append(pd.DataFrame(filtered))
df3 = pd.DataFrame()
for i in range (500):
    df3=pd.concat([df3,fil[i][0]], axis=1)
for i in range (500):
    df3=pd.concat([df3,fil[i][1]], axis=1)
for i in range (500):
    df3=pd.concat([df3,fil[i][2]], axis=1)
from itertools import product

```



```
L = ['Theta F','Theta N', 'Theta O','Theta S', 'Theta Z','Alpha  
F','Alpha N', 'Alpha O','Alpha S', 'Alpha Z','Beta F','Beta N',  
'Beta O','Beta S', 'Beta Z']  
k = 100  
L2 = range(1, k+1)  
df3.columns= [x+ str(y) for x,y in list(product(L,L2))]
```

**Lampiran 6.** Program Python untuk Ekstraksi Fitur Sinyal EEG

```

import pandas as pd
import numpy as np
a = df3.as_matrix()
b=dict()

#Fitur Energi
energies = []
for column in range(0,1500):
    b[i]=df3.iloc[:,column].isnull().sum()
    energies.append(sum(a[0:a.shape[0]-b[i]:,column] ** 2))
energy = pd.DataFrame(energies, columns=['Energy'])

#Fitur Maksimum
Maksimum = []
for column in range(0,1500):
    Maksimum.append(max(a[:,column]))
Maksimum=pd.DataFrame(Maksimum,
columns=['Maksimum'])

#Fitur Minimum
Minimum = []
for column in range(0,1500):
    Minimum.append(min(a[:,column]))
Minimum=pd.DataFrame(Minimum, columns=['Minimum'])

#Fitur Variance
Variance = []
for column in range(0,1500):
    b[i]=df3.iloc[:,column].isnull().sum()
    Variance.append(np.var(a[0:a.shape[0]-b[i]:,column]))

Variance = pd.DataFrame(Variance, columns=['Variance'])

```

```
#Fitur Entropy
result=dict()
entr=dict()
for i in range (0,1500):
    result[i] = df3.iloc[:,i]
    entropy.append(sum([(xi ** 2)*np.log10(xi**2) for xi in
result[i]]))
entropy=pd.DataFrame([entropy])
entropy=np.transpose(entropy)
entropy.columns=['entropy']
```

**Lampiran 7.** Program Python untuk Pembagian Data *Training - Testing* dan Pemilihan K-Neighbors Terbaik

```

y = np.squeeze(np.asarray(y1))
X = np.squeeze(np.asarray(df5))
from sklearn.neighbors import KNeighborsClassifier
neighbors = np.arange(1,50)
n=5
from sklearn.model_selection import StratifiedKFold
kf=StratifiedKFold(n_splits=n, shuffle=True, random_state=5)
total=dict()
ac=dict()
se=dict()
sp=dict()
cm=dict()
for i,k in enumerate(neighbors):
    cm[i]=[]
    total[i]=[]
    ac[i]=[]
    se[i]=[]
    sp=[]
    knn = KNeighborsClassifier(n_neighbors=k)
    for train_index, test_index in kf.split(X,y):
        X_train, X_test = X[train_index], X[test_index]
        y_train, y_test = y[train_index], y[test_index]
        knn.fit(X_train, y_train)
        y_pred = knn.predict(X_test)
        cm[i].append((confusion_matrix(y_test,
y_pred)).astype(float))
    for i in range (len (neighbors)):
        total[i]=[]
        ac[i]=[]
        se[i]=[]
        sp[i]=[]
        for j in range (n):

```

```

total[i].append(sum(sum(cm[i][j])))
ac[i].append((cm[i][j][0,0]+cm[i][j][1,1])/total[i][j])
se[i].append(cm[i][j][0,0]/(cm[i][j][0,0]+cm[i][j][0,1]))
sp[i].append(cm[i][j][1,1]/(cm[i][j][1,0]+cm[i][j][1,1]))
sensitiviti = np.empty(len(neighbors))
spesifisiti = np.empty(len(neighbors))
akurasi = np.empty(len(neighbors))
for i in range (len (neighbors)):
    akurasi[i]=np.mean(ac[i])
    spesifisiti[i]=np.mean(se[i])
    sensitiviti[i]=np.mean(sp[i])

#Plot nilai k-neighbor terbaik untuk masing-masing masing
performansi
plt.plot(neighbors, akurasi, label='akurasi')
plt.plot(neighbors, spesifisiti, label='spesifisitas')
plt.plot(neighbors, sensitiviti, label='sensitivitas')
plt.legend()
plt.xlabel('Nilai k')
plt.ylabel('Performansi')
plt.savefig('Pemilihan nilai k',dpi=300)
plt.show()

```

**Lampiran 8.** Program Python untuk Klasifikasi KNN dengan Nilai K terbaik

```

from sklearn.naive_bayes import GaussianNB
knn = KNeighborsClassifier(n_neighbors=3)
n=5 #jumlah fold
from sklearn.model_selection import StratifiedKFold
kf=StratifiedKFold(n_splits=n, random_state=1000)

cm=[]
total=[]
ac=[]
se=[]
sp=[]

for train_index, test_index in kf.split(X,y):
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    knn.fit(X_train, y_train)
    y_pred = knn.predict(X_test)
    cm.append((confusion_matrix(y_test, y_pred)).astype(float))
for j in range (n):
    total.append(sum(sum(cm[j])))
    ac.append((cm[j][0,0]+cm[j][1,1])/total[j])
    se.append(cm[j][0,0]/(cm[j][0,0]+cm[j][0,1]))
    sp.append(cm[j][1,1]/(cm[j][1,0]+cm[j][1,1]))
akurasi=np.mean(ac)
spesifisiti=np.mean(se)
sensitiviti=np.mean(sp)
#performansi masing masing fold->csv
df_k3=pd.DataFrame()
test=dict()
for j in range (3):
    test[j]=[]
for i in range (n):

```

```
test[0].append(ac[i])
test[1].append(se[i])
test[2].append(sp[i])
for i in range (3):
    df_k3=pd.concat([df_k3,pd.DataFrame(test[i])],axis=1)
df_k3.columns=['Akuras','Sensitivitas','Spesitifitas']
df_k3
```

*(Halaman ini sengaja dikosongkan)*



## BIODATA PENULIS



Penulis dengan nama lengkap Carianto Hottua Sihombing dilahirkan di Doloksanggul pada 14 Februari 1996. Penulis menempuh pendidikan formal di SDN 173397 Doloksanggul, SMPN 2 Doloksanggul, dan SMAN 1 Doloksanggul. Kemudian penulis diterima sebagai Mahasiswa Departemen Statistika ITS melalui jalur SBMPTN pada tahun 2014. Selama masa perkuliahan, penulis aktif di berbagai kepanitiaan salah satunya adalah kamjin CERITA 2016 yang merupakan *event* seminar nasional yang diadakan oleh HIMASTA-ITS. Selain itu, penulis juga aktif dalam organisasi yang menaungi salah satu Divisi HIMASTA-ITS yaitu PSt (Professional Statistics) HIMASTA-ITS sebagai Tim Ahli pada periode 2016/2017. Selama menjalani masa perkuliahan, penulis diberi kesempatan untuk menjadi semifinalis lomba matematika yang diadakan oleh Himpunan Mahasiswa Jurusan Matematika Vektor Universitas Negeri Malang pada periode 2016 dan 2017. Selain itu, penulis juga berkesempatan menjadi salah satu wakil ITS dalam bidang matematika pada lomba Olimpiade Nasional Matematika dan Ilmu Pengetahuan Alam (ONMIPA) pada tahun 2016. Apabila pembaca ingin memberi kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini, dapat menghubungi penulis melalui email [cariantohs@gmail.com](mailto:cariantohs@gmail.com) atau nomor telepon 082276350983.